1988-06

A methodology for validation of high resolution combat models

Coville, Michael Paul

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/23305

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

A METHODOLOGY FOR VALIDATION OF HIGH
RESOLUTION COMBAT MODELS

by

Michael Paul Coville

June 1988

Thesis Advisor: Samuel H. Parry

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b RESTRICTIVE MARKINGS

2a SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release: distribution is unlimited.

4 PERFORMING ORGANIZATION REPORT NUMBERS(S)

5 MONITORING ORGANIZATION REPORT NUMBER(S)

6a NAME OF PERFORMING ORGANIZATION ORGANIZATION
Naval Postgraduate School

6b OFFICE SYMBOL (If applicable)
39

7a NAME OF MONITORING ORGANIZATION
Naval Postgraduate School

6c ADDRESS (City, State, and ZIP Code)
Monterey, California 93943 - 5000

7b ADDRESS (City, State, and ZIP Code)
Monterey, California 93943 - 5000

8a NAME OF FUNDING / SPONSORING ORGANIZATION

8b OFFICE SYMBOL (If applicable)

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c ADDRESS (City, State, and ZIP Code)

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
A METHODOLOGY FOR VALIDATION OF HIGH RESOLUTION COMBAT MODELS

12 PERSONAL AUTHOR(S)
Coville, Michael P.

13a TYPE OF REPORT
Masters Thesis

13b TIME COVERED
FROM ________ TO ________

14 DATE OF REPORT (Year, Month, Day)
1988 June

15 PAGE COUNT
102

16 SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17 COSATI CODES

FIELD	GROUP	SUB - GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Validation, Combat Models, Simulation, National Training Center

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Senior officers in the United States Army have a high degree of confidence that National Training Center simulated combat results are representative, under similar circuatances, of actual combat. A validation methodology for high resolution combat models, primarily based on data from the National Training Center, is the focus of this thesis. The validation methodology, where appropriate, translates confidence in National Training Center realism, to confidence in the model. Theoretical issure, existing mehtodologies, and the impact of model purpose are considered in this research. The final product is a validation methodology that makes use of a realistic representation of combat, automatically updates validation criteria to account for changes in weapons and tactics, and is responsive to the purpose for which the model was designed

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED / UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL
Samuel H. Parry

22b TELEPHONE (Include Area Code)
(408) 646-2779

22c OFFICE SYMBOL
Code 55Py

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

Approved for public release:distribution unlimited.

A Methodology for Validation
of
High Resolution Combat Models

by

Michael Paul Coville
Captain, United States Army
B.S., United States Military Academy, 1978

Submitted in partial fulfillment of the
requirements for the degrees of

MASTER OF SCIENCE IN SYSTEMS TECHNOLOGY
(Command, Control, and Communications)

and

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 1988

ABSTRACT

Senior officers in the United States Army have a high degree of confidence that National Training Center simulated combat results are representative, under similar circumstances, of actual combat. A validation methodology for high resolution combat models, primarily based on data acquired from the National Training Center, is the focus of this thesis. The validation methodology, where appropriate, translates confidence in National Training Center realism, to confidence in the combat model. Theoretical issues, existing methodologies, and the impact of model purpose are considered in this research. The final product is a validation methodology that makes use of a realistic representation of combat, automatically updates validation criteria to account for changes in weapons and tactics, and is responsive to the purpose for which the model was designed.

TABLE OF CONTENTS

I. INTRODUCTION

A. WHAT IS VALIDATION

High resolution combat simulations are used across a broad spectrum of military activities. One sees their use and influence in the training of military forces, in the development of weapon systems, in the analysis of operational plans, in resource allocation planning, and in the development of doctrine and tactics. However, this widespread use is not without criticism and concern. The basis of this concern is a question of confidence. What is the appropriate level of confidence a decisionmaker should or should not have in the results of a combat simulation? This concern generalizes to include the question of relative confidence between differing simulations. The question of confidence is of extreme importance. Whether or not a combat simulation will be used at all depends on the level of confidence a decisionmaker has in it.

The issue of confidence begins with the type of problems that simulation is used to address. Combat simulations are generally used to address "squishy" problems because other methods of analysis are inadequate. The "squishiness" of a problem refers to how well it can be defined quantitatively; the more "squishy", the less well defined [Ref.1, p.43].

> If the real world problem we choose to solve by means of simulation were simple, and the solution set straightforward, we would not waste our time modeling. It is the complex, multidisciplined problems with convoluted solution sets that we attempt to solve by modeling and simulation. [Ref.2, p.21]

Since defining the problem is difficult, interpretable, and open to argument, the structure, processes, and results of the simulation become questionable. Numerous questions are generated. "Are the assumptions and transformations of the model correct?", "Can we believe what the model is telling us?", "Is the model useful?", "Why is this simulation better

or worse than another?", and "Is the simulation a good representation of reality?". These types of questions were summed up in 1968 by Dr. W. Fain, Chairman of the 1968 Warfare Model Verification Conference:

> The question is, are the models good abstractions and do they relate to the real world. [Ref.3, p.4]

This question, while well posed, still presents some problems. What is "good" and what is the "real world"? Each person may define these terms somewhat differently, and with each differing definition there may be a different answer to the same question. Little significant progress has been made in addressing the decisionmaker's major concern of confidence.

Objective consideration of the question posed by Dr. Fain, as well as consideration of "good" and "real world" as they relate to simulation, falls within the realm of validation. Validation concerns itself with changing the trust one has in a model from a trust based on faith to a trust based on objective analysis [Ref.4, p.298]. Validation is the process associated with this change.

B. PURPOSE AND SCOPE OF THESIS

The National Training Center at Fort Irwin, California enjoys a strong reputation for representing combat in a very realistic fashion. Moreover, senior officers in the United States Army have a high degree of confidence that the results from the National Training Center are representative of results that would be achieved, under similar circumstances, in actual combat. This confidence is the basis upon which the results and lessons learned of the National Training Center impact on policy decision made by these officers.

The purpose of this thesis is to develop a validation methodology that, where appropriate, translates confidence in the National Training Center to confidence in the model under investigation. The validation of high resolution combat models against a standard source of comparative criteria

would have beneficial effects for the Army. It would provide an objective alternative to advocacy as the primary source of model validation within the Army. It would also provide a method of standardizing the comparison of models. Finally, a methodology based on a realistic representation of combat would strengthen the Army's ability to cull out those model that are inappropriate representations of combat.

The theoretical issues associated with the process of validation are outlined and discussed in Chapter 2. These issues are extremely important because, far from having only philosophical impacts, they also significantly affect the practical matters of model validation. They bound one's ability to conduct validation, but also provide direction by highlighting the important issues that any validation methodology must address.

In Chapter 3, consideration is given to the existing methodological approaches. Naylor and Finger's multi-stage approach is shown to be most comprehensive, but fails to provide proper consideration to model purpose. The purposes of high resolution combat models are discussed, as well as their impact on the model validation process. Based on this analysis, the multi-stage approach is modified to incorporate model purpose into the methodology.

With a general methodology established, the requirement for an acceptable reference system is addressed in Chapter 4. The reference system is the measure of reality against which a model is judged during the validation process. Three candidates, expert opinion, historical combat data, and exercise/test data are analyzed with respect to their individual advantages and disadvantages. This analysis results in a "best" choice for use as a reference system in the validation process.

An analysis of the National Training Center as a reference system and the refinement of the general methodology to make use of NTC data are the topics of the

final two chapters. The final product is a validation methodology that makes use of the most realistic representation of combat, automatically updates validation criteria to account for changes in weapons and tactics, and is responsive to the purpose for which the model was designed.

II. THEORETICAL ISSUES

Validation of combat simulations, and models in general, continues to cause much pause within the modeling community. The problems associated with validation have not eased over time. Even as the arsenal of data collection methods, statistical techniques, and other tools, through which we attack the validation problem, have grown, the continuous desire for increased model detail has offset these gains. Theoretical considerations are many and have been with us since Aristotle's time.

One of the major underlying problems is one of definition. As described in the introduction, the question of "reality" comes immediately into play. Defining reality establishes the standard against which the simulation is compared. Without such a standard, validation cannot be accomplished.

Besides the difficult task of defining reality, there are three other significant theoretical issues.

> "The Teleological Problem"--How a model by its nature formulates an explicit cause-and-effect relationship that excludes other proximate or remote causes.
>
> "The Epistemological Problem"--How the "truth" of any model is always provisional and dubious.
>
> "The Uncertainty Principle"--How the very act of formulating or exercising a model distorts the reality we seek to represent. [Ref 4, p.303]

These four theoretical problem areas will significantly impact any methodological approach to validation and therefore deserve individual consideration.

A. DEFINING REALITY

The problem of defining reality has plagued philosophers and scientists for centuries. The difficulty is that reality is a fleeting essence, changing from minute to minute, and argued by some to exist only as an idea in the minds of men. In the context of validation and combat simulations, and from

a more practical viewpoint, the best one can hope for is a reference system that will generate a consensus of use upon which further considerations can be based.

> The real {reference}[1] system is nothing more than a source of potentially acquirable data. At any point in time we will have acquired only a finite subset of this data from what is an infinite set or universe. In general, the real system is (or will become) a source of behavioral data consisting of time based trajectories of input, state and output variables. [Ref.5, p.574]

Many reference systems have been proposed. Each has its strengths, its weaknesses, its advocates, and its enemies. During the 1968 Center for Naval Analysis conference on the topic of validation, actual combat was proposed as the appropriate standard of reality. Combat data, while appealing because of their source, exhibit significant weaknesses in accuracy and completeness. These weaknesses are, of course, reasonable considering combat has a purpose quite different from that of providing data to beleaguered modelers.

Other proposed reference systems include tests and exercises, and the judgement of experts. Test and exercise data, while offering significant gains in accuracy over combat data, carry the burden of being measures of abstractions of actual combat. Thus, even though the accuracy of the measurement may have increased, the reference system itself is now only a second order representation of actual combat. Often the greatest insights can be gained through critical examination by those who are knowledgeable of and experienced with combat. Human nature, however, is a stumbling block for effective use of "experts" in establishing the standard for reality. Generalization from personal experience is often hampered by the parochial aspects of the experience, and by the perceptual biases of the individual. Another fear associated with the use of "experts" is that the "experts" are often the clients for whom the simulation is being developed. [Ref 4, p.302]

---

[1]. { } authors addition.

B. TELEOLOGICAL PROBLEM

Teleology refers to explaining events in terms of final causes. Every model is a representation of a set of cause and effect relationships. In the broadest sense, they are the input-output transformations of the model, and in a more micro sense they are the interrelationships established within the model. The events of the world, including war and combat, are part of a continuous, dynamic stream of existence, interwoven into a fine fabric that details finer and finer level of cause and effect relationships. The teleological problem is that every model, of necessity, must start the representation at a particular level within this fabric of life. In making this choice of a starting point certain cause and effect relationships are excluded from representation. The level of choice is identified by the assumptions and inputs upon which the model operates. The teleological problem, as it relates to combat modeling, was particularly well illustrated by Wayne Hughes.

> Teleology is the study of final causes. A model always asserts a certain cause and effect, even when it has sophisticated feedback loops.... We presume a cause when we write inputs.... The model not merely asserts presumed first cause, but circumscribes for its user the world of admissible causes.
>
> Consider a warfare example: Why did Lee lose at Gettysburg? Historians may take as proximate cause the ill-conceived charge of Pickett on the third day. Or possibly Meade's artillery, massed in the center.
>
> As causes "once removed," there was Meade's astute tactical leadership and Lee's uncharacteristic tactical error. But few historians stop there. The cause was "really" J.E.B. Stuart's absence, so that Lee fought blind. Or the earlier death of his stalwart Stonewall Jackson.
>
> Deeper still, it was simply the inevitability that sooner or later the odds would catch up with Lee, and his daring battlefield tactics would overextend him. The fundamental cause, therefore, was the union's greater mobilization base. Lee was impelled by a sense of urgency, knowing that time was against him. Thus, what historians may call a tactical blunder was Lee's last-gasp gamble, a gamble made with a thoroughgoing appreciation of the true odds against breaking through the center.
>
> None of the "causes" above is unimportant, and the list is by no means exhaustive. One could add the Union quartermasters' efficiency ("logistics dominate war"), the motivating reasons why the soldiers fought tenaciously, etc.

> All the "causes" contributed to the effect: the Confederates lost the battle. Any model of it will emphasize some things and deemphasize others, even to the point of exclusion. Whether the model is the analyst's simulation or the historian's description, it circumscribes the event with some set of cause-and- effect relationships. Any model, even the most ambitious, is vulnerable on grounds of sufficiency -- its omission of the n-th order "cause-of-a-cause-of-a-cause...." [Ref.4, p.304]

As Hughes points out, every model has a particular level of circumspection, which establishes the teleological limitations associated with the simulation. Attempts at validation of the simulation, then, are bounded by the limitations introduced through consideration of the teleological problem.

C. EPISTEMOLOGICAL PROBLEM

Epistemology, the theory of knowledge, concerns the many diverse issues associated with the human ability to "know".

> The questions which it investigates are those such as the character of knowledge itself and the relation between it and belief; the validity and reliability of our claims to knowledge of the external world through sense perception; the propriety of claims of knowledge beyond the limits of sense perception; our use of general concepts and of general words; and the presuppositions required for our use of memory and by our claims to recognize objects or kind of object as being the same as what we have met before. [Ref.6, p.419]

Different subsets of these questions have been considered the most important and have received the most attention at various time in history. In the twentieth century epistemology has "mainly concerned itself with questions of knowability of the external world as accessible to empirical observation for the verification of hypotheses. [Ref.6, p.249] Validation is strictly tied to epistemology in so much as it is a process that leads to the acceptance or rejection of certain claims based on "knowledge" of the real system under consideration. In fact, every validation methodology is based on one or more epistemological approaches to gaining and evaluating "knowledge" of the real world.

One example of an epistemological approach might be to base knowledge on what one can sense and measure of the real

world. The claim might be that knowledge gained in this fashion is obviously a true representation of reality. This claim requires closer examination. How can one be sure that one's senses and measuring instruments are providing an accurate representation of reality? Looking at a stick in water one might perceive it to be curved, but upon removing it from the water it is straight. If one could not remove the stick from the water, would it be straight or curved, and how could one substantiate either claim of knowledge? When looking for the truths of combat, how can one "know" when truth is observed or when "fog of war" still clouds perception? Shopehauer poses the problem in this fashion.

> no knowledge of the sun but only of the eye that sees the sun, and no knowledge of the land but only of the hand that feels the earth [Ref.7, p.347]

Even from a more practical point of view, it is easily seen that any knowledge gained in the manner is conditional upon the accuracy of the method of measurement. While there may be a true length associated with a particular rope the bounds of human ability to access that truth may preclude ever "knowing" it. Knowledge gained in this fashion is both conditional and associated with a particular level of uncertainty.

The impact of this is that given an empirically well defined reference system, and good agreement with the results of a simulation, one still may not logically conclude that the simulation is validated. Any claim of validation must be caveated with the limitations of the empirical approach.

Other approaches exist, but all fall short of adequately addressing the various issues associated with epistemology. However, from the many discourses on the many approaches, two tenuous points of consensus fall out. The first is that human knowledge, and the laws and theories based on that knowledge are never complete. The second is that an unavoidable characteristic of human knowledge is uncertainty.

Thus, within the realm of validation of combat simulations the epistemological problem can be stated as two questions:

1. Given a particular reference system (reality), what are acceptable methods of claiming knowledge of the system,how certain can one be about the knowledge gained,and what are the conditional limitations of the knowledge.

2. Given the lack of total knowledge of a system, and uncertainty associated with the available knowledge, by what standard or standards is the simulation compared to the reference system.

Due to the close relationship of validation and epistemology, approaches to validation deal primarily with answering the two questions stated above. Different methodological approaches for dealing with these two issues and others are considered in the following chapter.

D. UNCERTAINTY PRINCIPLE

Formulated by Werner Heisenberg in 1927, the uncertainty principle, while born to the science of physics, has had a significant impact on a great many fields of intellectual pursuit.

> It is to be emphasized that in observing a system it is necessary to exchange energy and momentum with it. This exchange alters the original properties of the system. The resulting lack of precision with which these properties can be measured is the crux of the uncertainty principle. [Ref.6, p.487]

Within the context of combat, application of the uncertainty principle to human behavior is of much greater consequence than its impact on the physical properties of the data collected.

Consider an observer/data collector on the battlefield. His presence and, more often than not, his purpose will be known to the leaders involved in the engagements he is observing. Even with the extreme pressure of life and death at hand, human nature will exact a price. The presence of the observer will affect the actions and decisions of the participants of the battle. In each leader's mind will be the hint that his decisions and actions will be chronicled for later review and analysis. So there may be a little more

bravado when concern is called for, a little softheartedness when hard decisions need to be made, or a little less risk taking than victory demands. Consider also the observer himself and what actions he may take if he is facing the possibility of death. Is it reasonable to expect the observer not to pick up a rifle and fight when his life is threatened?

Even if a human observer is not present, the act of measuring combat may affect the process one is trying to measure.

> as, for example, in World War II aerial bombing when some crews refused to drop bombs in certain unfavorable conditions after bomb cameras were installed in their planes because the combat film was used in a scoring system associated with efforts to improve the modeling of bombing accuracy [Ref.8, p.309]

While this effect can never be countered in total, every care must be taken to minimize changes to the reference system that are caused by trying to measure it.

E. SUMMARY

Consideration of these theoretical issues begins to shed light on the extreme difficulty of the validation process. It can now clearly be seen that a formal "proof" of a simulations replication of reality is an impossibility. Analysis of these theoretical issues supports the position that validation is something short of a "proof" and is not inherently a question that can be answered simply yes or no. While a "proof" is unavailable, these theoretical issues do not preclude the establishment of a reasonable level of confidence that the simulation adequately represents reality. In fact, they provide direction as to what needs to be done and limitations on what actually can be done.

The teleological problem and the uncertainty principle place bounds on what can be done. The first sets a lower bound on the claim to validation. Simulations represent cause-and-effect relationships down to a specific level, and validation of the simulation can only be claimed within the

domain established by that bound. This bound should be established prior to the initiation of any validation attempt.

The uncertainty principle precludes 100% validation, even within the bounds established by teleological considerations. As data is measured and collected on a particular reference system, careful and diligent efforts should be made to minimize the impact of these actions. The observed impacts as well as expected impacts should be tracked and reported as the validation process continues. The impact of changes of human behavior because of observation/measurement may be subsequently bounded through an a fortiori analysis.

These two issues are adequately addressed through tying the scope of the validation effort to the scope of the model, and through explicit treatment of the impact of measuring the reference system.

The remaining two issues,defining reality and the epistemological problem, require deeper consideration of the practical aspects of validation. Each of these issues is addressed in detail in the next two chapters, and their consideration establishes the framework for the development of a validation methodology incorporating National Training Center Data.

II. METHODOLOGICAL APPROACHES

A. EXISTING APPROACHES

1. Rationalism

The philosophy of rationalism is based on the idea that there exists some unquestionable truths "not themselves open to empirical verification or general appeal to objective experience." [Ref.9, p.612] The term synthetic a priori was coined by Immanuel Kant to describe these types of "truths." In his book Urban Dynamics, Forrester's urban model is based on a rationalistic approach which he defends in this fashion.

> Much of the behavior of systems rests on relationships and interactions that are believed, and probably correctly so, to be important but that for a long time will evade quantitative measure. Unless we take our best estimates of these relationships and include them in a system model, we are in fact saying that they make no difference and can be omitted. It is far more serious to omit a relationship believed to be important than to include it at a low level accuracy that fits the plausible range of uncertainty. [Ref.10, p.144]

The idea is to identify the unquestionable premises and test the logical development of the model from those premises. If the premises can be accepted and the logical development proves sound, the model is considered valid.

The problem with validation under this approach is that there is a significant difficulty in explicitly stating all of the "unquestionable" premises. Even if this could be achieved, rarely would a consensus on the "unquestionability" of the stated premises be possible.

2. Empiricism

This philosophy is diametrically opposed to that of rationalism. Empiricists fault rationalism for not basing model assumptions on empirical data, and lacking this, argue that models based on rationalism are meaningless and not representative of reality. Naylor and Finger present the objections this way.

> Although the construction and analysis of a simulation model, the validity of which has not been ascertained by empirical observation, may prove to be of interest for expository or pedagogical purposes (eg. to illustrate particular simulation techniques) such a model contributes nothing to the understanding of the system being simulated. [Ref.11, p.B-92]

Reichenbach goes even further, arguing that synthetic a priori simply do not exist.

> Scientific philosophy .... refuses to accept any knowledge of the laws of the physical world as absolutely certain. Neither the individual occurrences, nor the laws controlling them can be stated with certainty. The principles of logic and mathematics represent the only domain in which certainty is attainable; but these principles are analytic and empty. Certainty is inseparable from emptiness; there is no synthetic a priori. [Ref.12, p.304]

Empiricism requires that validity be established by testing assumptions on the basis of empirical data. While the problem with validation under rationalism was one of consensus, for empiricism it is primarily one of data. It is often extremely hard, especially for combat, to gather data that is acceptable for use in the empirical testing process.

3. Positive Economics

An objection to both the previous approaches was presented by Milton Friedman in his book Essays in Positive Economics. He argued that testing model assumptions was the wrong approach and that the true test of a model's validity rests in its predictive ability.

> The difficulty in the social sciences of getting new evidence for this class of phenomena and of judging its conformity with the implications of the hypothesis makes it tempting to suppose that other, more readily available, evidence is equally relevant to the validity of the hypothesis-- to suppose that hypotheses have not only "implications" but "assumptions" and that the conformity of these "assumptions " to reality is a valid test of the validity of the hypothesis different from or additional to the test by implications. This widely held view is fundamentally wrong and productive of much mischief. [Ref.13, p.445]

If the model consistently produces results that are born out in the real world, how important is it that the structures and processes underlying the model be congruent with those of the real world? The approach of positive economics considers these isomorphic requirements irrelevant.

If the behavior of the simulation's dependent variables are consistently and accurately predicted (at least better than any other existing model), then positive economics classifies the simulation as valid. After all, the "answer" is what the simulation is all about.

There are two approaches to testing the predictive ability of the simulation. The first deals with the ability to reproduce historical outputs given the same inputs, and is referred to as retrospective prediction. The second method deals with forecasting future events based on a specific set of inputs, and is referred to as prospective prediction. Validation through prospective prediction is the stronger test, however, this approach is not possible for combat simulations.

Critics of this approach, while agreeing that the predictive ability of a simulation is important, contend that it is in no way sufficient for validation of the simulation. While predictive ability is appealing, it is not appealing to falsify the structure and processes of reality, to whatever extent necessary, to make the "answers come out right." Furthermore, without an understanding of the structure and processes of the system under investigation, how can one know what real world changes will, at some unknown time, invalidate the predictive ability of the simulation. These problems are illustrated by a simple story.

> There was a student doing fractions, and he wrote down 16/64-at least the teacher wrote it down -- and the student cancelled out the sixes and got one quarter. And someone else objected, and the teacher said: "What's wrong? He got the right answer didn't he?" [Ref.3, p.54]

The teacher validated the mathematical model of the solution process based on the student's results but the problems with this approach are obvious.

4. Multi-Stage Validation

Originally coined as "multi-stage verification"[2], Naylor and Finger proposed this approach in 1976 as a method particularly well suited to validation of simular models.

> This approach to verification is a three-stage procedure incorporating the methodology of rationalism, empiricism, and positive economics. Multi-stage verification implies that each of the aforementioned methodological positions is a necessary procedure for validating simulation experiments but that neither of them is a sufficient procedure for the problem of verification. [Ref.11, B-95]

The first stage of this approach incorporates the rationalist methodology, but weakens the conclusiveness of tests applied. Naylor and Finger argue that the initial set of hypotheses upon which the simulation is based are found essentially through a search for Kant's "synthetic a priori" Given a particular real world system to be simulated, there are an infinite number of hypotheses that might be forwarded to explain its structure and processes. It would be impossible to empirically test each one as the method for selecting the best subset upon which to base the simulation. Only through the application of prior knowledge, past research, existing theory, and general observation of and familiarity with the real system, can this set of hypotheses be initially chosen. Any hypothesis that is questionable after careful scrutiny of this nature should be excluded from inclusion in the set of fundamental hypotheses. This test of "reasonableness" is an application of the rationalist approach. This process is commonly referred to as establishing face validity.

It is apparent, though, that experience with and knowledge of a system changes overtime. Thus what seemed reasonable one day may prove false the next, and conversely, what was unacceptable may be shown sound. This indicates that the test of reasonableness is temporal, and should be

[2] The terms verification and validation have both been used to describe the process of comparing a model to the real world. Validation dominates recent use in describing this process.

applied again and again as significant changes in the level of knowledge of the system occur. Naylor and Finger quote Reichenbach in this regard.

> Like the scientist, the scientific philosopher can do nothing but look for his best posits. But that is what he can do; and he is willing to do it with the perseverance, the self-criticism, and the readiness for new attempts which are indispensable for scientific work. If error is corrected whenever it is recognized as such, the path of error is the path of truth. [Ref.12, p.326]

Naylor and Finger break from the rationalistic approach at this point, rejecting the idea that these basic hypotheses require no further attempt at validation; "we merely submit these postulates as a tentative hypothesis about the behavior of the system." [Ref.11, p.B-96] This initial set of hypotheses is then used as input for the second stage of this validation approach.

The second stage incorporates the empiricist approach, and examines the set of fundamental hypotheses further. The hypotheses submitted from stage one are subjected to statistical tests based on real world data. Statistical theory, with respect to estimation and hypothesis testing, provides the basis for this stage of the validation process. Empirical testing, however, may not be possible. There may be some hypotheses for which there is no real world data available, or for which statistical tools are inadequate. One has two choices concerning hypotheses of this nature. The first is to simply reject the hypothesis, but this approach carries the burden of continuing the search for an acceptable hypothesis upon which to base the model. The second choice is to continue with the hypothesis in a "suspect" state. This is acceptable because there is no explicit proof that the hypothesis is wrong, but requires additional vigilance with regards to the impacts of this hypothesis. While the first is the more conservative approach, the costs associated with the reestablishment of the fundamental hypotheses may be prohibitive.

The third stage of this validation approach is to examine the predictive ability of the simulation. With only a narrow exception, Naylor and Finger argue that "the purpose of a simulation experiment is to predict some aspect of reality." [Ref.11, p.B-96] Thus it is that this final validation effort has a significant impact on convincing the user that the model does what it is supposed to. This stage of testing is done by comparing the input-output transformations of the simulation with those observed in the reference system. The methods by which this comparison may be made are quite varied. There are highly technical mathematical methods, such as spectral analysis, and behavioral methods such as "turing tests".

Naylor and Finger's multi-stage approach has been attacked on the grounds that it fails to give adequate consideration to the purpose of the simulation. This approach uses prediction as the only purpose of simulations, and while possibly true at one time, this certainly is not the case today. Simulations are used to instruct, evaluate policy alternatives, and develop theory as well as to predict output values. The multi-stage approach combines the strengths of the three previous approaches well, but is lacking in its explicit consideration of the possible impacts of the purpose of the simulation.

5. Absolute Pragmatist

This approach developed to a large extent in response to the multi-stage approach's failure to consider model purpose. It focuses on the simulation, much like positive economics did, as a black box. While positive economics viewed prediction as the only purpose of simulation, the absolute pragmatist approach broadens the horizon of uses. This approach argues that each simulation is developed for a purpose and it is the ability to successfully accomplish that purpose that establishes validity.

> We propose that the criterion of usefulness of the model be adopted as the key to its validation, thereby shifting the emphasis from a conception of its abstract truth or falsity to the question whether the errors in the model render it too weak to serve the intended purpose. [Ref.14, p.B-105]

The usefulness of a model has an easily arguable place in the validation process. If the model does not serve its purpose, it will not be used no matter how many other validation tests it may have passed. Showing that a model serves its intended purpose is the "bottom line" for decisionmakers. If the decisionmaker has no confidence in the model, it essentially does not exist.

Critics argue, as in the case of positive economics that while this criteria is applicable, it is not sufficient for validation. The question remains one of knowing the provisional qualities of the model, and when, based on input changes, the model is no longer valid.

B. IMPACT OF MODEL PURPOSE

Naylor and Finger present a comprehensive approach to the process of validation with the exception of their failure to address the implications of simulation purposes other than prediction. This section addresses the primary uses of high resolution combat simulations and the impact that these different uses have on the validation process. The intent is, in particular, to examine the effects of simulation purpose on the validation process, and to determine whether the multi stage approach is still appropriate with respect to purposes other than prediction.

1. Reproduction of a Real System

Reproduction of the real system is done to gain insight into its operation, and to predict the behavior of the system under particular conditions. As argued by Naylor and Finger, this is the purpose of most simulations. In cases where this is not the primary purpose, meeting reproducibility criteria generally assures the simulation is adequate for its primary purpose.

The criteria for validation, in this case, is how well the simulation replicates the selected reference system. Limitations of resources, time (for development and for running the simulation), money, and data limit the accuracy to which the modeler can replicate the real system. The question is whether the simulation's level of isomorphism is adequate to predict system behavior and provide understanding of system behavior. The comparison, to gauge this accuracy, is generally accomplished through empirical testing.

2. Comparison of Courses of Action

This is a primary use of high resolution combat simulations. Comparisons of courses of action are undertaken to make decisions on weapon procurement, tactics, and force/weapon mix strategies. The decisionmaker wants information on the relative value of the alternatives available to him. In this case the actual values of the simulation are not as important as the accurate representation of the relative differences between competing alternatives. The simulation must provide a discernable representation of these differences, and the accuracy of the representation must be such that appropriate decisions can be made. When the decisionmaker only needs to know which decision is best, representation of the relative differences also becomes unimportant, and proper ordering of the alternatives is all that is needed.

The validity of the model is determined by its ability to appropriately represent the real system to the level required by the decision under consideration. While this requirement is less rigorous than strict replication criteria, reproducibility is still the dominant criteria. If the simulation accurately replicates the real system, then the relative values of outputs for different courses of action will also be representative of the real system.

3. Instruction

When the simulation is used to instruct or train, the paramount consideration is that the model impart to the student proper lessons about the real system under study. In other words, the simulation must not teach the student inappropriate responses, or provide the student with false insights. Consider a simulation developed to teach a lieutenant the proper method of employment of his platoon in clearing a minefield. The simulation might represent losses associated with this action as stochastic in nature. If the probabilities are accurately developed from historical data, the predicted outcomes may be very representative of the long term losses associated with clearing minefields. However, if the lieutenant learns that losses are a product of chance, the model failed in its purpose. Training is conducted in "snapshots", and if the "snapshot" does not reinforce the proper lesson, it does more harm than good. Another outcome of the stated situation might be that the stochasticly produced losses associated with a poorer course of action may be lower than losses associated with a superior method. This disparity would correct itself in the long run, but the lieutenant is learning from the "snapshot" of reality that the simulation has produced. In this case the lieutenant may have again learned the wrong lesson. When models are used for instruction, the need seems to be for the model to operate in a fashion that consistently provides outcomes that reward application of currently approved doctrine and tactics. For specific purposes (teaching that attacking the enemy flank is better than a frontal attack) certain model parameters might be somewhat exaggerated to drive the lesson home. The validity criteria for this type of simulation is no longer strictly tied to replication of the real system.

> the validity criteria have shifted from the observable universe to the cognitive and affective systems of those individuals whom the operating model is intended to instruct. [Ref.15, p.219]

If a different simulation was developed for each different lesson to be taught then manipulating parameters to support these lessons would be appropriate. The costs associated with this type of training approach would be enormous, and therefore the requirement is for simulations that can be used to teach the broad range of skills and techniques associated with combat. Due to the extreme interdependence of the processes and entities involved in combat, adjusting one parameter to support a particular lesson generally detracts from the ability to teach other lessons. The need is for an appropriate middle ground, and this middle ground is accurate replication of the real system. While the lesson that the student learns is still of the greatest importance, replication of the real system supports the broadest range of lessons, and provides realism as the student is learning.

4. Examination of Non-existent Universes

A working prototype of a particular weapon system has not yet been built, yet combat simulations are used to examine the effects of its use in particular combat scenarios. Tactical nuclear weapons have not been used against US forces in Germany, yet simulations are used to address this potential engagement. Simulations are used again and again in the development of contingency plans for scenarios that may never occur. Combat simulations used in this way are examining "non-existent universes." Validation of simulations with this purpose is extremely difficult. In this case there exists no observable universe that offers reference points by which one can check the veracity of the assumptions associated with those yet to occur events.

Two types of future systems are examined by combat simulations, those that are the result of revolution and those that are the result of evolution. The first, indicating a future state substantially different than the present, occurs primarily when examining highly futuristic

weapons or extreme catastrophic conditions. The time and effort spent in this area is less, due to the lower probability of occurrence, than investigation of the second choice.

Investigation of future states that are the result of evolutionary variations of the present is even more dominant when considering high resolution combat simulations. Future states resulting from evolutionary change are those states that are reached through incremental change in the structure and processes of the present state. Considering that "the most powerful determinant of what will happen tomorrow is what is happening today" [Ref.16, p.122], comparison to the present state may provide some measure of the confidence that should be associated with the simulation. This comparison is reasonable because, in evolutionary development of future states, the incremental change affects only a small percentage of the existing present state hypotheses.

Even for evolutionary future states, the comparison of the future to the present becomes untenable when either one or both of two conditions exist. The first condition is a large time gap between the present and the future state under consideration. When the time difference is large, the evolutionary chain between the present and this particular future state becomes weaker and weaker. The longer away the future state is, the greater the permutations of event paths available for the future to have progressed along. The second is if the evolutionary changes occur over a significantly broad range of present day hypotheses. As the number of changed present day hypotheses grows, the basis of comparison between the present and the future once again weakens. The greater the number of changes the weaker the link between present and future. In fact, at some point the changes may, in sum, cause the future state to be more representative of revolutionary change than of evolutionary change. In considering either of these two problem areas the

establishment of what is too large a time gap and what is too many changes is subjective and judgmental. The more conservative the restrictions on time and change, the stronger the comparison is as a method of establishing confidence in the simulation.

In general, the criteria available for validation of simulations of this nature are logical consistency and reliability [Ref.15, p.219]. When the domain of consideration is limited to high resolution simulations and consequently to evolutionary future states, comparison of the simulation hypotheses and outcomes to the present state is an appropriate method of approaching validation.

C. REVISED APPROACH

While the purposes described in the sections above are not exhaustive, they represent the majority of uses of high resolution combat simulations. In each case model purpose has affected the criteria of the validation process. Referring back to the original question, "are models good abstractions and do they relate to the real world," the impact of model purpose is on how the model relates to the real world. What relation is represented and to what extent is the relation represented are the considerations governed by the model purpose. This is seen in the varied criteria for validation. For system reproduction the criteria is direct replication; for comparison of COA it is tempered replication; for instruction it is the effect on student cognitive processes; and for non-existent universes it is logical consistency and reliability.

Within each of these somewhat varied validation criteria there does exist a common thread, and that thread is replication of an existing reference system. In the first two cases it is explicitly stated, and in the last two cases replication becomes a practical, useful criteria by default. In so much as the multi-stage approach explicitly treats replication as a criteria, its applicability in each case is

supported. However, there is a provisional requirement. Since model purpose refines the implementation of the criteria, the multi-stage approach must account for this refinement.

A method of incorporating model purpose into the multi-stage validation process is to use model purpose to establish the initial criteria for validation. The criteria would be consistent across models of the same purpose but would be allowed to change when model purpose differed. Thus, model purpose would be used to divide models into classes, within which the validation criteria would be the same. A revised approach could then be described as follows.

1. Define model purpose and establish a framework of validation criteria based on the purpose.
2. Establish face validity.
3. Empirically test model hypotheses.
4. Empirically test the model's predictive abilities.

IV. THE REFERENCE SYSTEM

Accepting the previously addressed theoretical problems associated with defining reality, consideration is now turned to the more practical issue of establishing the best reference system to represent reality. This reference system will represent the baseline from which the validity of a simulation will be judged.

Characteristics of a "good" reference system are accuracy of representation, detail of representation, and accuracy of measurement. The first addresses the ability of the reference system to capture the causal relationships of the real system. The second characteristic concerns itself with the level of technological detail the reference system provides to the modeler. The final characteristic concerns itself with the measurement accuracy the reference system offers of the interactions and effects of the represented relationships.

As previously mentioned, there are three reference systems most often proposed for the validation process. These are expert opinion, historical combat data, and exercise/test data. Each of these will be addressed and assessed in regards to their advantages and disadvantages as a reference system.

A. EXPERT OPINION

Expert opinion consists of the views, perceptions, instinct, and acquired knowledge of those who have been and are closely associated with the system under study. Depending on the system under study "expert" status can be gained through experience with the system, or through academic study of the system. In the case of combat, it is a mix of both of these elements that characterizes an expert. The most qualified expert is one who has an experience base

that has been continually and extensively expanded through academic endeavor. The application of expert opinion as a reference system would involve the use of expert opinion to identify the correctness of hypotheses associated with particular combat processes. A consensus of some type would need to be generated and documented. This reference system, while consensus could be difficult, could be updated periodically as the climate of combat is perceived to change over time.

1. Advantages

Those who have experienced combat and have studied the various aspects of war have particular insights into the actual relationships and structures of combat. These insights cannot be replicated with numerical descriptions of combat. They are based on a conscious and subconscious understanding of the intrinsic relationships of combat. To a large extent they represent the behavioral content of combat. Weapon systems, in an inert or controlled environment, can be adequately described through mathematical representation of their characteristics. This is not the case when man, and consequently human behavior is involved. How does the inclusion of man, who has the ability to gather and process information and change his behavior accordingly, affect system performance? How do the intangibles; leadership, morale, group cohesiveness, and courage, affect the relationships inherent in the system? Attempts at the quantification of human behavior in combat have not met with much success [Ref.1, p.32]. Until progress in this area occurs, the major source of information about the effects of these variables will be expert opinion.

A second advantage of expert opinion is its ability to present a holistic interpretation of the processes and structures of combat. In general, the application of scientific methodology to the study of combat divides combat into component parts, examines the simpler parts, and then

rebuilds the system. This process overlooks the intrinsic relationships between various components of the combat. One of the most important concepts not captured by this approach is that of synergism. The expert can provide this view of combat. He can identify those hidden interactions that make the sum of the parts greater than the whole.

The formulation and interpretation of "squishy" problems are unavoidably judgmental and are inherently connected. Thus, if an experienced professional officer, speaking of a particular hypothesis, says "This doesn't make sense and here's why," one would be ill-advised to ignore his comments.

2. **Disadvantages**

Just as the advantages of expert opinion revolved around human behavior considerations, so do the disadvantages. The way each person is brought up, the inherent position of the individual, and the goal orientation of an individual affects the way he views the world and the way he records what he views. Different people identify different issues as being the most relevant to the events they are viewing or experiencing. Consider, for example, a combat engagement experienced by three soldiers. One is a lieutenant, another is a sergeant, and the last is a private. Each will be sensitive to certain aspects of his environment and even though all three went through essentially the same experience, the differences in their accounts of the experience may be large. A more macro example of perceptual bias is captured in the phrase "The winner gets to write the history books." Recounts of the progress of the events of World War II, the causal relationships between those events, and the relative importance of different events ,receive different emphasis depending on whether the basis of knowledge is from an American, a Russian, or a German. The question that becomes relevant in this case is which view best represents the reality of what occurred. The

perceptual bias may be undetectable when experts of similar backgrounds, culture, and experience are providing the representation of reality.[3]

Related deficiencies in the use of expert opinion for a reference system are a lack of detail and quantitative accuracy. The human mind is limited in the amount of detail it can provide with regard to specific events. This lack of detail is usually caused by overflow in the short term memory during the event occurrence [Ref.17, p.646]. Thus while experts can provide a very realistic,insightful description of combat processes on a general scale, as the need for more detailed data grows, the experts falter. Human limitations in quantitative information processing also detracts from the effectiveness of expert opinion as a reference system. While one is generally willing to say which weapon is better than another, when asked for a number that describes how much better, answers come hesitantly. Wholistic reasoning is relatively easy for humans but quantitative, computational reasoning is much more difficult [Ref.17, p.645].

A less serious disadvantage is one of parochialism. Knowledge that an expert gains from experience is often local, and therefore provisional upon the circumstances and environment of the experience. The provisional aspects of the experience are often forgotten as the experience is translated to a broader scope. It is part of human nature to inductively transfer local experiences into general rules. When the number of local experiences is limited, as is the general case of combat experience, the generalization of personal experiences to general rules is hazardous. This disadvantage of expert opinion can never be completely overcome, but certain steps can be taken to minimize its impact. One is to give the problem an appropriate amount of

---

[3] A detailed list of cognitive biases is provided in Appendix A. These biases effect both perception and recall, and with the effects of short term memory impact heavily on human quantitative ability.

concern, and another is to limit the effect of parochialism by amalgamating the experiences from many sources.

A final disadvantage of expert opinion is in the relationship between the experts and the modeler. Particularly in the military, the experts are the clients of the modeler. The problem is, then, to what extent do modelers bend objectivity, and sound hypotheses to please their clients.

> --perhaps many of us poor analysts have yielded to the pressure of our customers and our friends, and we are discovering we are all members of the same club. We are all yielding to the pressures and modifying our work because it doesn't suit General So-and-So's intuition, so we've got to pull it back a little bit over here, and lo and behold, when we run something which is essentially a probablistic type solution, sure we get a number that lies between zero and one, and somebody else has too, and it probably lies where people want it to be. [Ref.3, p.110]

The experts that have experienced combat and have studied war extensively are one and the same as the senior military leaders who have dedicated their lives to military service. These senior military leaders are the decisionmakers that say "yea" or "nay" to the purchase and use of a particular simulation. When these decisionmakers also provide the basis of the reference system upon which the simulations are developed, the models reflect the predisposition of the decisonmakers and all the underlying motivations to which they are subject. These underlying motivations may be other than to provide a realistic, useful model.

B. HISTORICAL COMBAT DATA

It is again worthy to note that the strongest determinant of what will occur tomorrow is what is happening today. The present and the future are inextricably tied to the past. The only "real" data on "real" combat exists in the past tense. Historical combat data is that information that has been gathered from past conflicts.

The ordered collection and analysis of this type of information has only begun to be seriously addressed in recent years. One of the strongest proponents for the use of

historical data for validation, Col.(Ret) Trevor Dupuy, founded the Historical Evaluation and Research Organization. This organization is the only such organization in the U.S. pursuing this extensive cataloging and analysis of historical combat data [Ref.18, p.na]

The use of historical combat data in the validation process makes the validation question one of whether the retrospective fit of the simulation to the past is strong enough to warrant confidence in the simulation. Historical data would provide input and parameter values to the simulation. The simulation would then be run and the outputs of the simulation would be compared to the results of history.

1. **Advantages**

The first and obvious advantage of historical combat data is that it comes from actual conflict. All the "dirty" aspects of war are captured in this data. The impact of the actual level of troop training, the failure of communications, the imperfect execution of orders, the havoc weather plays on the best of plans, uncertain intelligence, and all the implications of less than perfect logistics are represented in this data. More importantly, this data is a true reflection of human involvement in the combat system. While most of the other factors might be adequately estimated in other ways, the implications of facing life threatening situations is still largely a mystery. Only "real" combat data is from situations where men actually faced the immediate prospect of losing their lives. This aspect of war cannot be duplicated in peacetime.

Another major advantage of historical data is that it provides one the opportunity to investigate the time independent principles of combat.

> Although new technology, more sophisticated armaments, and indeed the new geopolitical implications of major conflicts have demanded changes in the art of warfare, no one can afford to ignore what has been done in the past. Whatever the changes in methodology and tactical concepts, basic principles that have found their roots in the evolution of

> warfare itself remain very much the same. It is therefore from the sound knowledge of former battles, from the study of military thought, that one can refine one's judgement, develop one's skills, and have a basis for developing the new tactical concepts necessary to the modern battlefield. [Ref.19, p.vii]

One need only examine the writings of Sun Tzu, Saxe, Clausewitz, and Jomini to find evidence that these principles exist. Each of these men identified similar hypotheses regarding certain relationships in combat. The fact that they show up in the writings of men vastly separated by history argues for the existence of these time independent principles. Historical data is the only reasonable source for investigating time independent trends and subsequently refining and defining mathematical hypotheses that describe these trends.

2. **Disadvantages**

While combat data has an intuitive appeal for use as a reference system, it is,unfortunately, replete with shortcomings and pitfalls. The first revolves around the previously discussed issue of the purpose of combat. Its purpose is not to provide data for later analysis, and therefore, the participants primary concern is not with the collection and recording of such data. Combat data suffers extensively form both a lack of completeness and of accuracy. Even when good data is available on the output side of the conflict (ie. attrition, movement of frontlines,etc.), the input variables have never been well recorded [Ref.20, p.336]. These variables include the amount of ammo available, the actual orders issued, and many others. Another issue in the area of completeness is the one sidedness of the data collected. Data on the enemy, either input or output, is much harder to come by. Information on the size of the enemy force in any engagement, their tactical procedure, and their logistical status is often lost as soon as the battle is over. The enemy, as do we, take conscious steps to keep this type of information from becoming available. The result is that even when friendly data is

fairly complete, the historical data is not usable because the two sided aspect of conflict is not represented.

There are two primary sources of historical data: archives and official military histories. The National Archive data is spotty and requires great effort to extract. Figure 1 illustrates the incompleteness of these records. The availability of records about the 79th Division in late 1944 is depicted and the gaps are easily identifiable. [Ref.21, p.10]

79th DIVISION RECORDS IN THE NATIONAL ARCHIVES							
		1944					
Organization	Record	July	August	September	October	November	December
79th Infantry Division	G1 Journal		●	●	●		●
	G2 Journal	●	●				
	G3 Journal	●	●				
	G3 Journal File	●	●	●			
	G3 Maps	●	●				
	G4 Journal						
	Operations Report			●			
	Daily Diary	●	●	●	●	●	●
	Situation Reports	●	●	●			
Division Artillery	Journal						
	Journal File			●			
	Intelligence Summary				●	●	●
	Overlays		●		●		
313th Infantry Regiment	Journal		●	●	●	●	●
	S2 File						
	S3 File						
314th Infantry Regiment	Journal	●	●	●	●	●	●
	S2 File						
	S3 File						
315th Infantry Regiment	Journal	●	●	●	●	●	●
	S2 File	●	●	●	●	●	●
	S3 File	●	●	●	●	●	●

FIGURE 1

In the case of official military histories, published by the US Army, the accounts are rich with qualitative descriptions of the events of war, but they lack tables, graphs or appendices with quantitative data. Figure 2 notes the data available from a group of World War II Army histories. If the history systematically presented any data, the work received credit for data being present. The conclusion is straightforward, combat data is generally not complete enough to provide a reference system for the validation process. [Ref.21, p.12]

QUANTITATIVE DATA IN WORLD WAR II HISTORIES

Volume	Contained Some Attempt at Summarizing Data on:												
	Intelligence Estimates	Troop List		Personnel Counts		Weapon Counts		Attrition		Vehicle Counts		Supply Consumption	
		Enemy	US	Enemy	US	Enemy	US	Enemy	US	Enemy	US	US	Enemy
NORTHWEST AFRICA: SEIZING THE INITIATIVE IN THE WEST			●		●		●					●	●
SICILY AND THE SURRENDER OF ITALY			●										
BREAKOUT AND PURSUIT													
THE LORRAINE CAMPAIGN													
THE ARDENNES: BATTLE OF THE BULGE													
THE SIEGFRIED LINE CAMPAIGN													
LOGISTICAL SUPPORT OF THE ARMIES	■					■	■	■					
CARTWHEEL: THE REDUCTION OF RABAUL	●	●							●				
VICTORY IN PAPUA													
LEYTE: THE RETURN TO THE PHILIPPINES			●						●				
THE APPROACH TO THE PHILIPPINES								●	●				
TRIUMPH IN THE PHILIPPINES		●	●	●	●			●	●				
CAMPAIGN IN THE MARIANAS		●	●	●	●								
OKINAWA: THE LAST BATTLE		●	●	●	●			●	●				
THREE BATTLES: ARNAVILLE, ALTUZZO, AND SCHMIDT			●										

● Some Systematic Data Present ■ Not Applicable

FIGURE 2

A general lack of accuracy also is prevalent in historical combat data. One ill-fated example of this was the body counts of the Vietnam War. After an engagement, dead bodies of the enemy were counted and reported to higher headquarters. These reports were often best guesses rather than accurate reports of the dead. This occurred for a number of reasons. The enemy, when possible, took their dead with

them. Additionally, even when friendly forces were forced back and no opportunity to count dead existed, body count reports were required. Leaders on the ground reported counts that included estimates of those uncountable dead. There were a distinct pressures to report more liberal than conservative estimates of the body count. These pressures were based on the political uses of the reported numbers, and the anticipated effect poor numbers would have on one's career. These types of pressures, while never exactly the same will always be there to affect the accuracy of any data collected during actual combat. The incompleteness and inaccuracy issues cause combat data to be reported and subsequently used in the aggregate. This, of course, is not acceptable when considering high resolution combat models.

Another significant disadvantage of combat data is that it does deal with past conflicts. A criticism of the American military is that it constantly prepares to win the last war fought. This comment emphasizes the change that is associated with combat. War is a competitive sport that has on each side intelligent, clever, industrious, and resourceful players, namely men. Man processes past information and constantly attempts to change the environment, climate, and conduct of battle to give his particular side the advantage. This relative advantage shifts again and again over time, and each shift is a shift away from previous characterizations of conflict. In particular, new weapons, new tactics, and new political objectives change the characteristics of battle. One only needs to review the effect of tanks in World War II to see evidence of the change in the character of battle.

As time passes, the gulf between past conflict and present day conflict increases. Since most of American conventional combat data is from World War II and the Korean War, this gulf is significant. In fact, the characteristics, interactions, and results of present day conflict may well be

outside the domain of possibilities established by this historical data.

Finally, the same problems of perceptual bias as described in the case of expert opinion, are present in historical data. These biases are impossible to deal with. While with expert opinion, the experts could be interrogated to establish the presence of bias, no such opportunity exists with historical data. Historical data rests on unalterable pages of print, most often without clues as to who, how, and under what conditions it was recorded. Thus, bias existing in historical data has a greater impact than if it exists in expert opinion.

C. EXERCISE/TEST DATA

Exercise/test data can be characterized by its three major sources. The most basic is technical engineering test data. This data establishes the pristine technical characteristics of weapon systems. Pristine is meant to imply than humans are not yet included in the domain of the weapon system, and environmental conditions are strictly controlled. This data is used to define the characteristic boundaries of weapon system performance. This data is useful in combat modeling only as starting baseline from which parameters and hypotheses can be further refined.

The second type of exercise/test data is from highly structured field experiments. In this case the data represents system performance when humans have been included in the system domain. The system performance now is a factor of hardware performance and human performance. Environmental factors are still highly controlled and it is generally the aim to establish system characteristics based on the interaction of humans with the hardware. Independent variables are changed incrementally to investigate and record their effects on system performance. Most often, these system performance characteristics are established while attempting to maintain human performance at an optimum. In

other words the stresses related with human performance under combat conditions are explicitly excluded from the exercise.

The final type of exercise/test data is from open form field exercises. These exercises are usually force on force and allow for as much realism as safety restrictions will permit. While some may be simulated,all aspects of combat are generally included in these exercises. Regular soldiers are used and human behavior and performance is allowed to take its normal course. Environmental factors are also uncontrolled. These exercises provide data on weapon system performance as it interacts with the other elements of the battlefield. Data is also collected on all other combat operating systems.[4]

1. Advantages

These data, as opposed to historical data, can generally be collected to any practical level of completeness and accuracy. Modern technology provides many methods for accomplishing this collection. The limiting factors in the completeness and accuracy of exercise data are resources and poor planning. Accuracy and completeness come at a price, as the desire for more complete and accurate information grows, the cost of acquiring that information grows manyfold. Often after an exercise is over, an analyst will bemoan the lack of a particular piece of information. In most cases, collection of this data could have been easily incorporated into the exercise plan, but poor planning precluded it. Given a reasonable amount of resources and proper planning, exercise/test data provides the most complete and accurate representation of the events that have occurred.

Objective data is also an important advantage of this approach. In general, one can explicitly eliminate most biases from the data collected. Much of the data collected

---

[4]. Current military doctrine defines seven combat operating systems: Intelligence, Maneuver, Fire Support, Air Defense, Mobility/Countermobility, Combat Service Support, and Command and Control.

is from instrumented sources and as such is less subject to bias than that collected by human sources.

Another significant advantage of exercise/test data is that it represents the current state of conflict. Current weapons are used and current tactics are employed. In most cases, current enemy capabilities and weapons systems are represented as closely as possible. Depending on the level of realism attained, this data has a closer relationship to future conflicts than does historical data.

A final advantage of exercise/test data is documentation. Records of who collected the data, how it was collected, and under what conditions it was collected provide greater insight as the data are analyzed and interpreted. Correct documentation also provides the opportunity for independent review and reduces the possibility of misuse of the data.

2. Disadvantages

The major disadvantage of exercise/test data is the lack of realism that often exists in these exercises. One unavoidable cause of unrealistic conditions is the requirement for safety restrictions that would not be in effect in time of war. It is not actual combat and therefore, soldiers are not put in uncontrolled high risk situations. Another factor that detracts from realism is participants knowledge that it is, in fact, an exercise. This knowledge removes many of the pressures associated with combat. Some of the pressures and stress are replaced by other stress inducing variables, but there is no doubt that human behavior in exercises is different than that in combat. Realism is also hampered by the devices and methods used to record the desired data. The devices for data measurement are often connected to or carried on weapon systems and may inadvertently change the operating characteristics of those weapons. These devices may also have an effect on the way the operator interacts with a particular system. The methods

of data collection may impose requirements on the participating organizations that alters their standard operating procedures. Thus, even though the data collected may be highly comprehensive and extremely accurate, it may reflect events and relationships that different from those that would actually exist in combat.

A by-product of participant knowledge and the lack of transparency in measurement devices and methodologies is gamesmanship. Gamesmanship describes the use of known artificialities of the exercise to bias the outcomes and processes of the exercise in one's favor. The exercise participants adjust their behavior to maximize the benefits offered by these artificialities. This adjustment of behavior is natural and expected of soldiers in a combat environment. The problem is that in combat they are adjusting their behavior based on changes in real world inputs, while in the exercise the adjustments are based on the artificialities of the exercise. Consequently, the events and processes observed are not reflective of reality but of the artificialities of the exercise.

D. SUMMARY

The question remains, "What is the best choice for a reference system in support of the validation process?" Figure 3 summarizes the analysis in terms of the stated characteristics of a good reference system.

Each option exhibits deficiencies in one area or another. The nature of these deficiencies make it difficult to evaluate them relative to each other. Another approach other than direct comparison may be taken to identify the option that will provide the best reference system. This approach is to examine the possibilities of eliminating the deficiencies currently attributed to each option. This examination may provide evidence that supports a particular choice.

REFERENCE SYSTEM SUMMARY

	REALISM	COMPLETENESS	ACCURACY
EXPERT OPINION	BETTER	WORST	WORST
HISTORICAL COMBAT DATA	BEST	BETTER	BETTER
EXERCISE/TEST DATA	WORST	BEST	BEST

Figure 3

Removing the deficiencies associated with expert opinion involves changing the mental characteristics of humans. It would require an increase in short term memory capacity, and a change in the way human beings process information. This is unlikely in the near future, if ever, and therefore precludes serious improvements of expert opinion as a reference system.

Alleviating the problems associated with historical combat data may be approached in two ways. The first approach involves locating historical data that has not yet been brought to light. This data might then, if extensive enough, increase the level of completeness and detail of existing historical data. Considering the amount of effort this would entail, this approach would only be reasonable if there existed large amounts of "undiscovered" historical

combat data. This is not likely and this approach offers little help in alleviating the deficiencies of historical combat data. The second approach is only mentioned for completeness and involves the requirement of a new conflict in which data of appropriate completeness and accuracy might be gathered.This approach is set aside without discussion due to its obvious detractors.

The problem with exercise /test data is one of realism. If there were methods to increase the level of realism in exercises, this data might prove worthwhile as a reference system. The limitations in achieving realism are primarily technological shortcomings and the obvious unwillingness to kill soldiers in exercises. Technological advancements are being made constantly, so the opportunity for eliminating lack of realism from exercise data does exist.

Of the three choices, only exercise/test data offers a reasonable approach to overcoming deficiencies associated with use as a reference system for the validation process. In fact, efforts to introduce more realism into training exercises has been a top priority in the Army for years. The next chapter examines the National Training Center as a source of detailed, realistic, and accurate data for use as a reference system for the validation process.

V. THE NTC AS A REFERENCE SYSTEM

It has long been the policy of the United States Army to train its personnel in the manner that they are expected to fight in combat. Paramount in this goal has been the continued effort to conduct this training under conditions that are as close to combat as possible. The replication of combat conditions include environmental conditions, the scenario, the enemy, and behavioral considerations (fear, stress, etc.).

Modern weapons and equipment have increased the tempo, lethality, and size of the modern battlefield. These changes have made it increasingly more difficult for the Army to ensure realism in home station training. The close proximity of civilian communities limit the use of aircraft, electronic warfare, live fire, smoke, and gas even though they are real world components of the modern battlefield. Land in these areas has competing uses, and the Army is hard pressed to establish large expanses of land for training. Additionally, home stations do not have the resources to maintain an "enemy" against which to train.

The culmination of efforts to overcome these ever increasing deficiencies was the development of the National Training Center (NTC). The NTC is located in the Mojave Desert at Fort Irwin, California. It encompasses 1000 square miles of rugged mountains , dried up lakes, and open desert [Ref.22, p.1]. The nearest civilian community is located 40 kilometers away. The NTC has the specific mission of providing realistic training to Battalion size units and below.

A. ESTABLISHING REALISM

Many factors are considered in providing a realistic environment for training at the NTC. Units deploy to the NTC

in the same fashion that they would deploy to actual combat. Their training is conducted over a period of fourteen days with little rest or respite from the environment of the Mojave Desert. About ten combat missions are conducted including live fire training and force on force engagements. The unit's higher headquarters as well as logistic, artillery, engineer, and air assets deploy with it to maintain realism in the command and control structure and the other functional systems of combat. The most important of these factors deserve further attention.

1. "Enemy" Force

An opposing force of two battalions, one armor and one infantry, is maintained at the NTC. These soldiers wear Soviet style uniforms and are trained in the methods and tactics the Soviets use in combat. Their training is updated periodically to ensure that the opposing force methods and tactics stay current with enemy doctrine and procedures. American vehicles and equipment have been visually modified to be extremely representative of their Soviet counterparts.

2. Maneuver

Choice of vehicle and unit speed, driving techniques, and vehicle formations are all at the discretion of the commanders, leaders, and soldiers involved in the exercise. The size of the training area ensures that mission boundaries are not artificially influenced by training area boundaries[5]. Responsibility for maintaining the force (safety considerations, etc.) are left to the discretion of the unit commanders as they would be in actual combat[6]. The bottom line is if that's the way one would maneuver in combat that's the way one maneuvers at the NTC.

---

[5] There is one small animal watering hole that is in the maneuver area but off limits. It is incorporated into the relevant missions by designating it as a contaminated area.

[6] The high personnel and vehicle accident rates are an unfortunate testament to the realism of the training.

3. **Command and Control**

Realism in command and control is maintained by the training unit receiving all its orders from its parent unit as it would during war. The Brigade receives plans and orders from a Division cell that the NTC maintains. It then processes those orders and provides plans and orders to the task force as it would during actual combat. Command and control internal to the task force is exactly as it would be in combat. The leaders are responsible for the operation and the welfare of their men, without interference from the control elements of the NTC.

4. **Weapon System Engagement Simulation**

A critical portion of the establishment of realism at the NTC is the use of the Multiple Integrated Laser Engagement System (MILES)[7] to simulate the realistic exchange of weapon fires on the battlefield. This system provides realistic simulation of the weapons employed (range, relative killing ability, times of flight, etc.) and allows the "killing" of soldiers in the exercise. MILES provides a transparent, event driven method of casualty assessment, which is integral to any realistic representation of combat.

5. **Mobility/Countermobility**

The employment and clearing of all types of obstacles is allowed. Unlike other training areas, particularly those overseas, the NTC has no restrictions on what can or cannot be done to the land. What is done is driven by the tactical requirements of the mission. In fact, soldiers are expected to "dig in" every opportunity they get. The unit commanders of both the friendly and opposing forces decide on obstacle emplacement. These obstacles, while increasing the potential for injury, receive no special markings and are real word limitations to maneuver.

---

[7] See Appendix B for a detailed description of this system.

6. **Nuclear and Chemical Effects**

Nuclear weapons are not often played at the NTC due to the catastrophic results of their employment on battalion size units and below. Use of a "nuke", even a small one, effectively stops play at the small unit high resolution level. Chemical use, on the other hand, is exercised significantly. CS[8] is the primary simulator for the various types of gases and chemicals that are expected to be used on the battlefield. Soldiers react realistically, donning gas masks, etc., to avoid the discomfort caused by the gas. Tied closely to the use of CS are chemical detection packets. These packets are exactly the same as the real ones, except that they have been treated to exhibit the presence of a particular chemical agent. In this fashion the chemical play at the NTC is made sensitive to the different types of chemicals that may be used on the modern battlefield. When participating personnel are attacked with chemical munitions, and "contaminated" they are not allowed to assume an unprotected posture without carrying out a decontamination process.

7. **Electronic Warfare**

The isolation of the NTC offers the advantage of realistic use of jammers and other electronic measures against friendly communications. It is known that the enemy plans extensive use of jamming, electronic deception, and communication intelligence collecting in the next conflict. The opposing force employs all these methods in its "battle" against the friendly task force. In one instance, the opposing force successfully tracked and jammed battalion command level communication through three frequency changes. This is representative of enemy capabilities and in a large part depends on the communication discipline of the participating unit.

---

[8] A gas that, when inhaled or when it comes in contact with eyes nose or mouth, causes much discomfort. Commonly used to disperse riots. No long term detrimental effects.

8. Close Air Support

The Air Force provides fixed wing aircraft for the close air support missions at the NTC. Participating units must request and plan for this support in accordance with standard combat procedure. Close air support is available to both the opposing and friendly forces and is used in both the live fire training as well as the force on force exercises.

9. Logistics

Logistics is a significant, some say overriding, factor of combat operations that is often excluded from exercises. At the NTC the logistics play is even more real than other functional systems because the logistic processes are in fact real. They are real in the sense that soldiers really don't eat, vehicles really do run out of gas,weapons don't have rounds for firing, and broken equipment stays broken if the logistic system is not operated correctly.

a. Medical

Each soldier carries a card that indicates, in the event that he is "shot", what his particular "wounds" are, or that he has been "killed". Soldiers' weapons are deactivated when they are shot and they may not return to the battle until they have been properly treated by appropriate personnel. Those soldiers with "wounds" that would require evacuation in real combat must be physically evacuated at the NTC. If the soldier's "wounds" cause "death", the medics and logistic personnel are required to process the "remains". Those soldiers "killed" are pooled to provide a source of personnel to simulate the replacement system. They are returned to their units as replacements after the appropriate procedures have been completed.

b. Food and Fuel

These are real commodities provided in the same fashion as they would be in combat. The enemy may interdict supply lines and deny these supplies to the friendly force. If proper requests, tactics, and linkups do not occur, no

special administrative action is taken to alleviate the problem. It is up to the commanders and leaders to ensure proper logistic support.

c. Ammunition

Live ammunition is supplied in requested amounts for the live fire range. This ammunition is exactly what would be used in war and supply is controlled, as in combat, by the Required Supply Rate and the Controlled Supply Rate. This precludes an unrealistic abundance of ammunition.

In the force on force engagements, two methods of ammunition resupply and accountability are used. With the exception of Stingers[9] and indirect fire weapons, the weapon systems have blank rounds that simulate the firing effects of the weapon. These blanks also activate the MILES devices on each weapon. If the rounds run out,the weapon will not fire; if the blank rounds are bad the weapon will not respond, and these rounds must be physically transported around the battlefield. Blank rounds coupled with MILES provides an extremely realistic representation of weapon system firing and interaction.

In the case of stingers and indirect fire weapons, Colonel Larry Word, a senior observer at the NTC for over three years, provides an explanatory example of the process.

> If the commander wants to use the battalion mortars to smoke and wants to build up his smoke capability, he requisitions ammunition. A piece of paper will say, "I am a box of 4.2 smoke." To resupply the 4.2 platoon, two five-tons[10] might come rolling up and all they have in the front seat are a stack of these cards. In actuality they would be loaded with 4.2 ammo. The paper ammo is put in the FDC. If he fires twelve rounds of smoke, our controllers pull twelve of those cards that say, "I represent so much ammunition." When the paper runs out, he has run out of ammunition and must request additional. It has to be hauled up in the appropriate manner. [Ref.23, p.19]

---

[9] Stingers are a relatively new man portable air defense missile system. The are fire and forget type missiles and as such cannot be adequately represented by MILES.

[10] These are standard military trucks with a five ton load capacity.

Controllers inspect vehicles periodically and confiscate simulated ammunition that, if the ammunition was real, could not physically be carried on the vehicle.

d. Maintenance

All repair parts supply, replacement vehicles, and maintenance activities are conducted as they would be in combat. If a vehicle becomes "damaged" or "destroyed" due to a MILES hit, damage assessment is done and the vehicle must go through the combat maintenance process before it will be returned to the battle. If this requires maintenance personnel then they must go to the vehicle and stay with the vehicle for amount of time that would have been required for actual repair. If the vehicle would have required evacuation, a recovery vehicle must move to its location and drive with it back to the maintenance area. These requirements ensure that the availability of vehicles and equipment is realistically simulated.

10. Summary

A final consideration is of the men who participate in the exercises at the NTC.

> The soldiers that participate in these exercises are those soldiers who are expected to engage in actual combat if the need should arise. They are the soldiers who are subject to fallacies in judgment, who are susceptible to an opposing force commander's guile, and who are capable of seizing the opportunity of the moment. In other words, the data from these scenarios will emulate to as high degree as possible the results expected from combat due to actual hostilities. [Ref.27, p.4]

With the inclusion of those who would, in time of war, actually be doing the fighting, the NTC data captures the impact of human performance.

The result of these efforts is a battle environment as close to real combat as technology and safety restrictions permit. A final consideration is of the men who participate in the exercises at the NTC. NTC represents "the most realistic engagement simulation and live fire Battalion task force tactical training available to a modern peacetime Army" [Ref.24, p.v]. The true impact of the realism at NTC is well

illustrated by an NCO that described his own experience at the NTC.

> They had never faced five to one odds; faced an enemy that would close at 20 kph and accept the losses; or tried to acquire targets buttoned up, in full NBC protective clothing, while under artillery and smoke. [Ref.25, p.20]

An important consideration to note is the high degree of confidence that senior military leaders have in the realism established at the NTC. As a result of this confidence, NTC results have a significant impact on many policy decisions throughout the Army. If high resolution simulations could be validated against this source of combat realism, it is reasonable to expect that adequate correlation between the two might earn some level of confidence for the simulation. To accomplish this, data is required. The methods of data collection at the NTC and the status of data availability and usefulness are addressed in the next sections of this chapter.

B. DATA COLLECTION SYSTEMS[11]

Systems currently active in collecting data at the NTC are the Instrumentation System, Observer/Controller Logs, and Communication Tapes. These collection systems offer a broad range of both qualitative and quantitative information. The operation and characteristics of each system is separately considered in the following sections.

1. The Instrumentation System

The instrumentation system consists of three subsystems:

1. The Range Data Measurement System (RDMS)
2. The Core Instrumentation System (CIS)
3. The Live Fire Subsystem (LFS)

From these instrumentation subsystems three types of data are collected: raw field data, manual input data, and derived

[11] Even though upgrades have been conceptualized and some implementation plans made, consideration has been limited to existing systems. Improvements can be expected to increase the reliability of the data collected.

data. The RDMS collects the raw field data, manual input data is recorded through the CIS, and derived data is developed from manipulation of either or both of the previous data types [Ref.26, p.2]. Figure 4 depicts the organization of the major elements of the RDMS and the CIS. It also illustrates the transfer of data among the system.

a. RDMS

The primary components of the RDMS are colloquially referred to as the "B unit" and the "A station". B units are transmit devices mounted on participating personnel and vehicles. A stations are receiving units located on hilltops throughout the NTC. The A stations gather data from the B units and retransmit the data to a computer for storage and analysis. The A stations act as a distributed network of data collection nodes, while the B units are the data producers of the system.

The B units are integrated with the MILES system and are a source of many types of data. A position location signal is one of the primary data elements transmitted. It is continually transmitted by the B unit and is periodically received by the A station. These signals are omni-directional and through receipt at multiple A stations, triangulation is performed to accurately locate each vehicle. The system updates vehicle positions every fifteen seconds. The B unit also transmits data pertinent to the operation of MILES. The B unit transmits the time of weapon firing, the type of weapon firing, and the specific vehicle to which the weapon belongs. The A station gathers this data and sends it to a central computer. Additionally, as sensors on a target register receipt of a laser pulse, the B unit will transmit the near miss, hit, or kill status of the shot, and the type of weapon that fired. Each B unit is registered to a specific vehicle. This allows the linking of data to each vehicle and allows the differentiation of friendly and

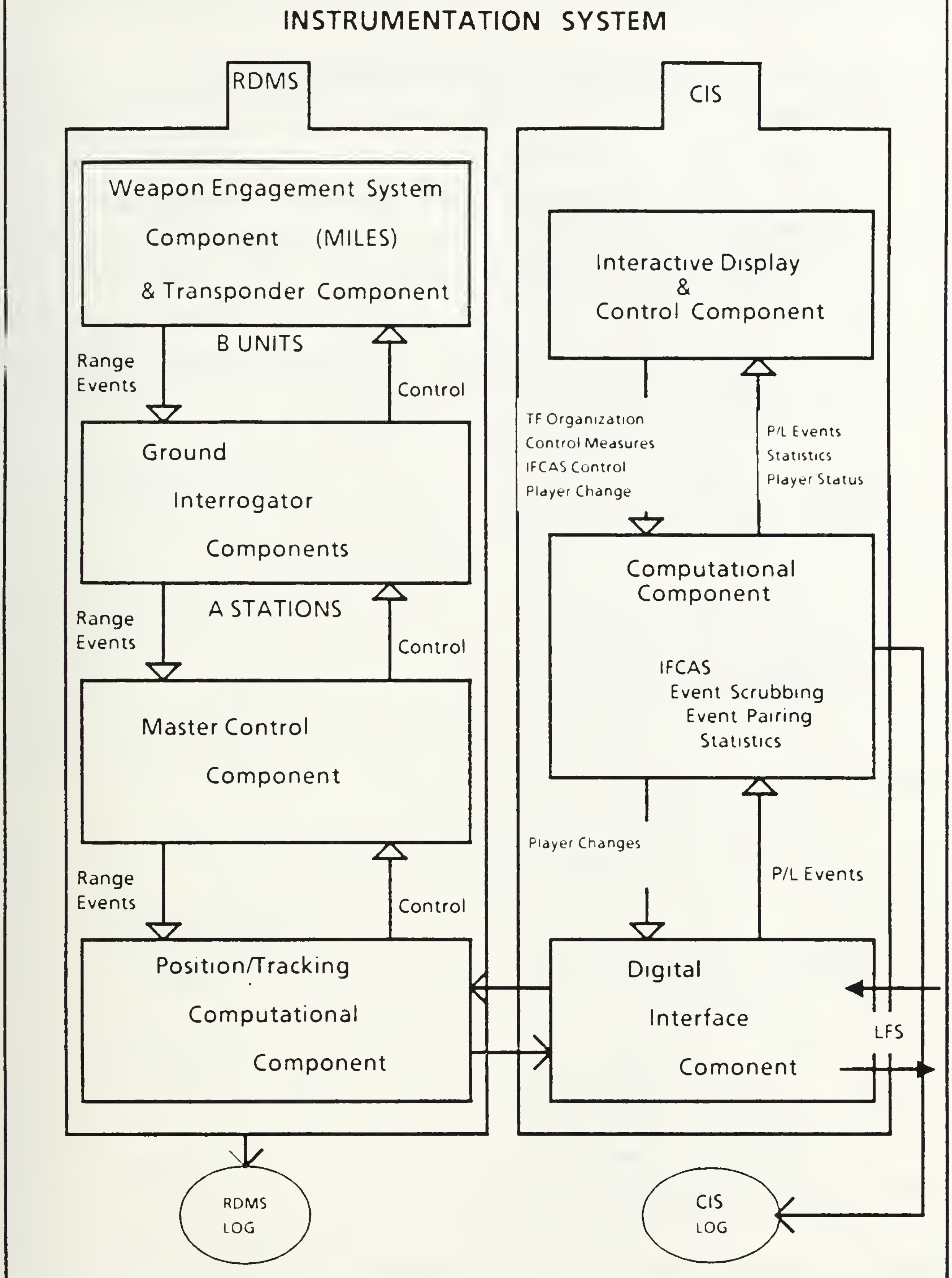
INSTRUMENTATION SYSTEM
RDMS
CIS
Weapon Engagement System
Component (MILES)
& Transponder Component
B UNITS
Range
Events
Control
Ground
Interrogator
Components
A STATIONS
Range
Events
Control
Master Control
Component
Range
Events
Control
Position/Tracking
Computational
Component
Interactive Display
&
Control Component
TF Organization
Control Measures
IFCAS Control
Player Change
P/L Events
Statistics
Player Status
Computational
Component
IFCAS
Event Scrubbing
Event Pairing
Statistics
Player Changes
P/L Events
Digital
Interface
Comonent
LFS
RDMS
LOG
CIS
LOG

Figure 4

opposing force elements. Figure 5 provides a synopsis of the data elements that are logged by the RDMS. [Ref.27, p.14]

b. CIS

The CIS, as the name implies, is the center of instrumentation action at the NTC. It interfaces with all of the other instrumentation systems and provides the computational support for real time data manipulation and feed back in support of the NTC training mission. The CIS provides interactive graphic displays with which controllers and analysts can "see" the battle develop.

The CIS logs data in real time and acts as the primary source of archival data. Initialization data regarding unit history and characteristics, as well as preplanned actions are inputted through the CIS. The CIS is responsible for the pairing of firing and target events from data input from the RDMS. This pairing is done through a time analysis matching of the input events. The CIS additionally provides real time control of the Live Fire Exercises, and receives input data from the Live Fire Subsystem. [Ref.28, p.57]

Another important function of the CIS is artillery casualty assessment. Indirect fires cannot be represented by MILES and require another method of realistically providing for their significant effects on the battlefield. The CIS receives firing data from the FDC's of the DS artillery supporting the battalion and the battalion's own mortars. It uses this information to run an internal simulation that projects projectile flight paths, and the burst location of the impacting rounds [Ref.28, p.58]. The system logs the event and the location of round impact. It then relays this location to the observer/controller in the field. The observer/controller first provides a visual and auditory cue of the incoming rounds: smoke and artillery burst simulators. He then assesses casualties based on the proximity of vehicles and personnel to round impact, the

DATA ELEMENTS OF THE RDMS LOG

Data Element	Description
Trigger Pull	Event received when a shot is fired by an instrumented weapon system. Event data consists of firer player number and weapon type.
Ammunition Remaining	Pair of events received immediately following trigger pull. Tens digit in former message, units in latter.
Laser Illumination	Event sent by target. This event is one of three differend kinds of codes, for HIT, NEAR MISS, and KILL.
Live Fire	There are four Live Fire events pased from the targets via RDMS. They are: target UP, target DOWN, HIT by ballistic projectile, and HIT by laser.
Communication	An event is sent by a player whenever the microphone key for either net is depressed or released. The message includes the net (1 or 2) and the action (on or off).
Position / Location	The Position/Location of each instrumented player is derived by RDMS software from raw signal data and logged.
Player Status	Player Status initialization and updates, which are entered from the CIS and transmitted to the RDMS are also logged. These data include the B unit player identification/weapon system assignment.

Figure 5

protective posture of the unit, and the type of round fired. Figure 6 provides a synopsis of the data elements logged by the CIS.

c. Live Fire Subsystem

The Live Fire Subsystem performs two primary missions. The first is to control the target array during the live fire exercise. The target array is developed to be a realistic representation of the formations used by Soviet forces. The second is to record event data form the live fire exercise and transmit this data to the CIS for processing and addition to the log.

The target array is made up of remote controlled vehicle and personnel targets. These targets are all outfitted with remote controlled fire effects devices. That is to say, when the target is displayed it simulates firing at the friendly forces through the use of certain flash and smoke devices. These devices are used to increase the realism of the target array. Additionally, the targets are cut to represent full size silhouettes of the vehicle they represent. The targets also have kill indicators that activate when the sensors of the target register a hit. The hit sensors register both ballistic and laser weapon engagements. The ballistic sensors have internal sensitivity settings that are set to maintain the appropriate hierarchial order of weapon systems on the battlefield. These settings ensure that targets representing tanks are not killed by small arms fire. The MILES sensors are used to capture the Dragon and Tow missile systems firing effects. These weapon systems are simulated by MILES due to the extremely destructive effect they would have on the target array. Destruction of the target array would require constant replacement which is fiscally prohibitive. [Ref.28, p.52]

Each target is equipped with a receiver transmitter over which it receives its commands and transmits the results of engagements. The control of the target array,

DATA ELEMENTS OF THE CIS LOG

Data Elements	Description
Background/ Documentation	History and mission name, start and end temes, mission type, exercise conditions, task force, and OPFOR organizations.
Unit/ Player Status Info	Status of individual players and/or units including: Instrumented / Not Instrumented Tracked / Not Tracked Position / Location
Fire Event (RDMS)	Event generated when a shot is fired by an instrumented weapon system. Should be identical to RDMS log with the exception of invalid events.
Pairing	Event generated when the laser sensors of an instrumented target system are illuminated and decoded into a valid message. If possible, target is paired with a firer.
Control Measures	Locations for control measures entered from IDCC. This includes control measure updates and mines.
Indirect Fire Casualty Assessment (IFCAS)	Fire mission number assessment of number of casualties inflicted.
Call Fire Missions	Call for previously planned inderect fire.
Commo	Player identification, radio net, and duration of commo messages longer than 55 seconds, should agree with RDMS log for those but all others are lost.

Figure 6

the collection of the target data, and the transmitting of this data to the CIS is accomplished by the range control system illustrated in Figure 7.

The computer at the center of the system is programmed to present a realistic target sequence over time. The computer keeps track of those targets that have been killed and does not present their subsequent representation to the friendly forces. This reinforces realism in that the participants of the exercise can see the effects of attrition as the enemy closes with them. The computer also records and stores target event data transmitted from the targets, and then relays it to the CIS.

There is a second portion of this system that monitors the actions of the friendly forces. Friendly weapon systems are fitted with interface devices that are keyed by the firing of the weapon. This device is connected to the position location system and together they provide firing event data and position data to the computer. The computer once again relays this information to the CIS for processing and evaluation. This data along with other pertinent data are logged in the CIS log for the live fire mission. [Ref.28, p.55]

2. Observer/Controller Logs

There are observer/controllers (OC's) watching every battle that occurs at the NTC. Their goal is to be unobtrusive, but accomplish these missions:

1. Enforce the rules of engagement
2. Assess indirect fire casualties
3. Implement indirect fire weapon effects cues
4. Record and communicate the results of friendly engagement simulation activities based on human observation. [Ref.24, p.11]

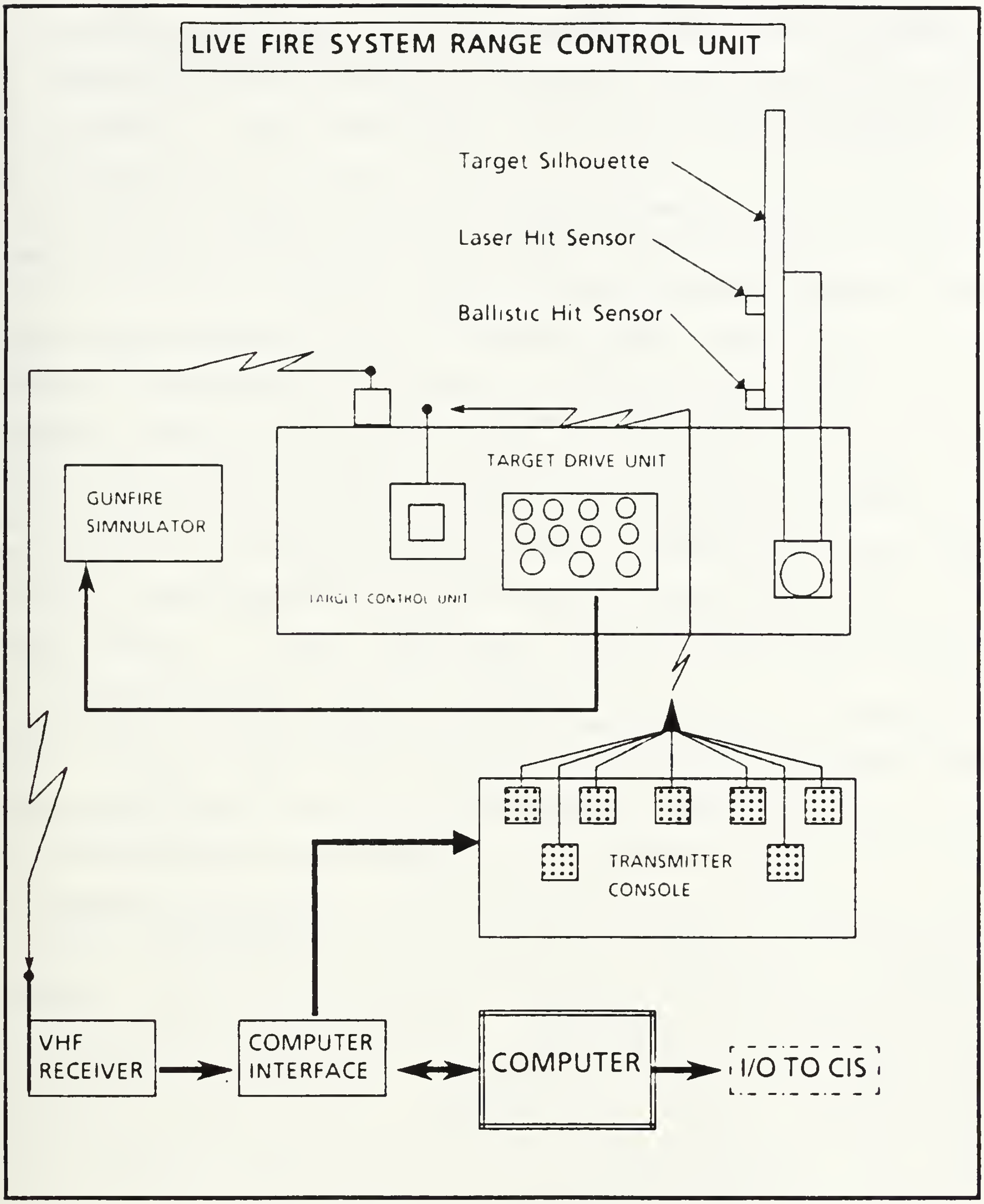

Figure 7[12]

To accomplish the last task, a number of manual records are kept, some of which get transferred to the database. These records supplement the data gathered by the instrumentation

---

12 This figure is adapted from the figure on page 54 of reference 28.

system. While the details of a kill are recorded in the CIS only if a pairing of firer to target is accomplished, the OC in the field gathers details on every kill. When a vehicle is killed in the field, MILES internally records the kill, as well as the weapon type of the killer. OC's collect and record information on all kills, paired or not, to complete the killing record. [Ref.24, p.8]

OC's also keep "notes" on the battles they observe. These notes are primarily used for discussion during the After Action Reviews, but are also significant sources of insightful information. The OC's can identify what facets of the battle were important or contributed most to the outcome. They can, for example, note that the soldiers of a particular unit were asleep due to exhaustion, and that as a result were caught offguard by the opposing force. This type of information is not available from the electronic data recorded by the instrumentation system, but is necessary to understand the "Why's and Wherefor's" of the battle.

A final type of manual log that is maintained is this artillery log. These logs are detailed records kept by the officers of the artillery Training Analysis Feedback Team (TAF) [Ref.24, p.8]. Indirect fires cannot be simulated by MILES and therefore event records of artillery firings are not automatically generated. While the event of firing and the impact point of the engagement are manually inputted into the computer system, the results of the casualty assessment are not. These results, along with other information, are maintained in the artillery logs.

3. <u>Communication Tapes</u>

The primary means of communication during tactical operations at the battalion level and below are tactical radios. The NTC maintains a 40 channel radio frequency monitoring system, that records transmissions over all nets used during the rotation. These tapes are an excellent source of descriptive detail and contextual information about

the battles recorded. Also, depending on the communication discipline of the administrative/logistic net, quantitative information on personnel and logistic operations is available.

Any attempt to use this data requires an extensive expenditure of manhours. The tapes, because of their nature, must be accessed sequentially, and the rate of information transfer is limited by auditory input capability. For a normal rotation of fourteen days these tapes represent 560 days of recordings. The tapes also provide no way of identifying signal overlap. This is the phenomenon of the closeness of hardware or frequencies causing bleedover from one channel to another. This bleedover would be recorded as normal transmissions. The tapes are also not time synchronized to allow comparison of same time communication on different channels. [Ref.29, p.8]

Due to the primarily qualitative data available from these tapes and the difficulty of extracting useful information from them, further consideration of their use as part of the reference system is discontinued.

C. DATA AVAILABILITY

Data supplied by the data collection systems are processed and then stored for future use in a NTC Research Database. This database is maintained by an element of the Army Research Institute at the Presidio of Monterey in California. The current database is a result of a recent (1987), extensive revision of the NTC database system. This revision was accomplished to eliminate excessive redundancies in the database and to enrich the content of the database in terms of the data that perspective users desired. The approach now used divides the database into two parts. The first is the tactical database and contains all digital data form the CIS and RDMS logs. The second part is the technical database that is developed to support specific research efforts.

TABLES OF THE NTC RESEARCH DATABASE

1. Mission Identification Table
2. Player State Initialization Table
3. Player State Update Table
4. Unit State Iinitalization Table
5. Unit State Update Table
6. Unit Type Table
7. Player/Vehicle/Weapon Code Table
8. Firing Event Table
9. Pairing Event Table
10. Communications Table
11. Ground Player Position Location Table
12. Air Player Position Location Table
13. IFCAS[1] Target Table
14. IFCAS Target Group Table
15. IFCAS Missions Fired Table
17. Minefield Casualties Table
18. Control Measure Table
19. Control Measure Add Table

Figure 8

The tactical database is composed of nineteen tables, and a separate one is generated for each mission. A list of these tables is shown in Figure 8, and a detailed listing of the data elements in each table is presented in Appendix C.

These tables and their associated data elements were chosen to allow for the inclusion of the maximum amount of information in a format that facilitates access for currently defined areas of research [Ref.30, p.2]. The database is implemented in an INGRESS relational database. This provides great capability for cross-referencing data, grouping data, and selecting data based on specified qualifiers.

D. DATA ANALYSIS

Thus far, the NTC has been found to simulate combat in a very realistic manner. Additionally, state of the art

technology is provided to collect the data produced, and it is subsequently stored in a fashion that supports research applications. The instrumentation system and other collection systems, however, are not perfect. The impact of the errors of the collection systems on the data collected is the subject of this section.

1. Digital Data

Though state of the art technology is utilized, equipment shortages, nature, and inherent characteristics of the collection hardware cause some corruption of the data. Major causes[13] of these distortions are:

1. Spurious radio frequency transmissions lead to erroneous events.
2. Internal "noise" in sensor systems that sometimes causes inaccurate pairing of events.
3. Normal hardware/electronic instrumentation problems leading to the loss or duplication of some events.
4. Coverage limitations (when vehicles enter arroyos, etc) cause loss of "track" which means no position/location data or event records during the time of loss of coverage.
5. Initialization inaccuracies occur when B units are not properly registered with the correct player and leads to improper assignment or the invalidation of events.
6. Equipment shortages that cause a number of the exercise participants to be uninstrumented and leading to the activities of some participants to not be electronically tracked.

The digital data that is most important to combat simulation, and most affected by these problems are the position/location data and the firing event data. Most studies investigating the impact of these irregularities have produced pessimistic results. In general, they find that the most severe problem is one of missing data [Ref.29, p.10]. Based on the missing data problem, arguments are presented about the non-usability of the NTC digital data for serious quantitative analysis. While no fault can be found with the numbers presented in these studies, the studies have some serious weaknesses. The data are examined in an aggregated fashion

---

[13] These problems are identified in Reference 26, p.2.

that treats all missing data as equal. In combat, as in most activities of life, certain aspects of, and certain participants in the activity have greater import than others. For example, consider the investigation of a firefight between two company size forces. The intent of the investigation is to gain insight into the combat between the two forces. Is it as important to know the location of the company supply truck as it is to know the locations of those participants actually fighting? This is not to say that supply activities are not important, but only that sometimes certain activities are not important to the question at hand. In other words, a 30% missing P/L data rate may be disheartening, but it is not a severe problem if the data are missing from elements that had no impact on the battle.

On the premise that different pieces of missing data might be more or less important than others, NTC P/L data were examined at a more micro level. P/L data were obtained from ARI for sample missions of sample rotations. The data were processed by a program that compared the P/L data to the task organization and thereby identified vehicles with missing data, vehicles with duplicate player numbers, and vehicles with bad P/L data. This information was then manually examined to identify the type of vehicles involved and there respective impact on the battle. Figure 9 depicts results of a typical examination.

The Blue Forces (exercise participants) had 50 combat elements that had no position /location data. This translates to a 35% missing data figure. If only this percentage is considered, the reliability of the data set is questionable and use of it might be extremely tenuous. Closer consideration of the participants without position/location data reveals some interesting insights.
The cavalry had a screening mission and were not directly involved in the battle. The artillery and the MLRS, while firing support missions for the task force, were located well

POSITION / LOCATION DATA

ROTATION 87-08
MA870806

Blue Forces

Total Forces = 141

No P/L data	Bad P/L data
12 - artillery	1 - M1
13 - cavalry	1 - Platoon manpack
8 - manpack stingers	1 - medic M113
7 - f/b	
5 - attack helos	
2 - radar	
1 - MLRS	
1 - M577 in CBT Trns	
1 - M113	

Red Forces

Total Forces = 252

No P/L data	Bad P/L data
18 - artillery	21 - BMP
16 - recon	12 - T-72
14 - manpack	1 - ZSU 23-4
14 - BMP	1 - 122mm HOW
5 - T-72	
4 - BMP (TOC)	
4 - helos (TOC)	
3 - f/b	
1 - T-72 (TOC)	
1 - ZSU 23-4	

Figure 9

to the rear of the battle area. The manpack stingers are man portable air defense missiles whose operators are assigned to subordinate units in the task force. The weapon is an area coverage weapon and as such the exact location of the weapons system is rarely required. The particular M577 missing data in this case is a vehicle that belongs to the personnel and logistic officer of the task force. As such it is the controlling element and travels with a group of vehicles known as the Combat Trains. The location of the Combat Trains is known from other vehicles in the group and the location of the M577 can thus be, albeit not exactly, established. The f/b entry represents the Air Force fighter/bombers that are allocated assets of the task force. These forces, obviously, do not continuously remain in the battle area but enter, deliver ordinance, and leave. Records of these point events can be established from flight records but the exact location of these fighter/bombers is not required throughout the battle. Thus the impact of the missing data has relatively quickly been whittled down to that of 8 vehicles. This represents a 5% missing data figure compared to the initial 35%. Figure 9 also indicates the minimal impact of bad P/L data.

While the numbers associated with the Red forces are significantly higher and represent a number of weapon types that would contribute to the battle, Red Force tactics minimize the impact of the missing data. The Red force operates in a very structured fashion and examination revealed that position data for BMPs and T-72s could be well established through examining the location of other vehicles in their units. For red vehicles of types comparative to the Blue Force, the arguments reducing or eliminating the impact of the missing data are the same. The only significantly different problem in the Red Forces is the missing data on their reconnaissance vehicles. These vehicles often play an important role in the development of the battle and because

of their mission, their positions cannot be established through comparative means. This problem can be overcome, with significant effort, only through extensive examination of all available sources of exercise information.

The conclusion from this closer examination of the impact of missing P/L data is that when considered in the context of the battle, the impact is minimal.

Examination of the firing event data produced similar results. The instrumentation system captures most but not all of the firing events that occur. This is evident when the computer records are compared to the O/C log of firing events. The computer records are consistently somewhat sparser than the O/C logs. Use of both of these sources provides comprehensive coverage of the firing events. Pairing of firer to target is much weaker. Common values associated with successful pairings are in the 5 to 10 percent range. If pairing does not occur the type weapon responsible for a kill is not recorded. Thus, while the connection between the firer and the target cannot always be established, the firing events, target events, and data on what type of weapon was responsible for kills is adequate in terms of accuracy and completeness.

Cursory examination of other digital data indicates that, in general, digital data collected by the NTC instrumentation system is, with some cross-referencing and scrubbing, suitable for use in the validation process.

2. Manual Data

This is data that are recorded manually by either the O/C's or members of the TAF. Some of this data is entered and maintained on the NTC Research Database and some is not. Two of the most significant problems with this type of data are:

1. Lack of standardization regarding observations recorded by the O/C's. Standardization would support quantification of information and statistical manipulation, permitting more concise interpretation of the results.

2. Large amounts of data requiring manual entry are missing. Much of this data would be useful in completing the picture of the battle. [Ref.29, p.9]

These deficiencies generally apply to the descriptive qualitative entries of the manual records. This portion of the manual records are important in establishing any irregularities associated with a particular battle. These entries support selection of "average" battles that do not exhibit extreme conditions or irregular circumstances. This descriptive, qualitative data is not useful as part of the comparative reference for the validation process because of the unidentified, yet unavoidable biases of the O/C's.

The tabular data that is manually recorded is of much more use. The collection format and the source of the data eliminate subjective biases from this data. Examples of this tabular data are the artillery logs of indirect fire events and the kill records of the O/C's. These records are of primary use in complementing and completing the data record established by the instrumentation system. There are limited problems with completeness but these problems can be avoided by choosing samples missions appropriately.

The use of the tabular data that is manually recorded with the digital instrumentation data provides sufficient usable data to establish the NTC as a reference system.

E. SUMMARY

Examination of the National Training Center as a candidate for use as a reference system in the validation process has met with encouraging results. The NTC offers close to real representation of combat, and provides significant amounts of usable data about the events and activities that occur. The NTC data is continually reflective of current weapon technology and of the current tactics and doctrine of both American and enemy forces. The NTC has overcome the most serious problem with using exercise

data as a reference system, that of realism, and as such offers the best choice for a reference system in the process of validation.

VI. VALIDATION METHODOLOGY

The previous chapters of this thesis were devoted to establishing a foundation for the development of a methodology for the validation of high resolution combat simulations. After identifying the theoretical problems associated with validation, attention was given to choosing a general approach to the validation issue. Naylor and Finger's multi-stage approach was adjusted to account for the impact of model purpose on the validation process. This revised multi-stage approach is the basis for the development of a more refined methodology of validation. The remaining requirement for completing the foundation was a reference system against which the combat simulation could be compared. National Training Center Data was evaluated as the best choice for a reference system.

The revised multi-stage approach consisted of four steps, as illustrated in Figure 10.

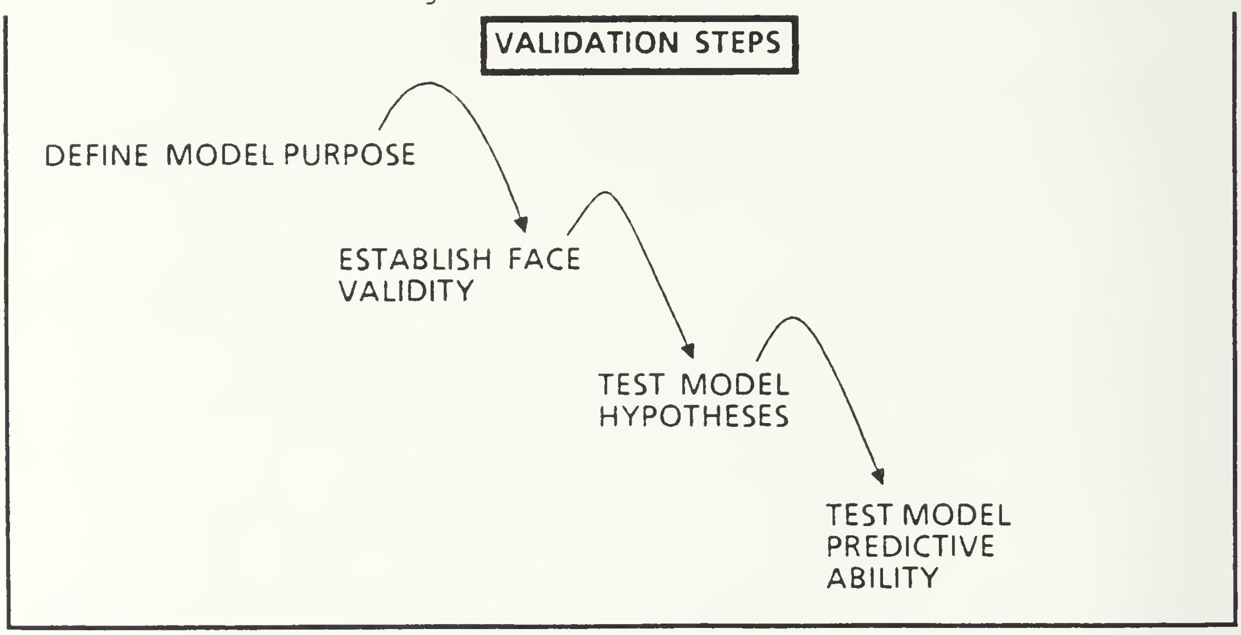

Figure 10

Each of these steps will be considered and expanded upon in this chapter.

A. DEFINING THE PURPOSE

As previously discussed the purpose for which the simulation is developed has an impact on the validation process. Specifically it affects the criteria against which a simulation should be judged. Model purpose also affects the stringency to which the evaluation criteria are applied.

Model purpose establishes, from the set of available evaluation criteria, the subset of criteria that are applicable to the validation of that model. For given purpose W let w_i represent the selection variable for a particular criterion, i. If the criterion is not applicable to the validation of models designed for purpose W, then the value of the variable will be zero. If the criteria, i, is applicable it will be assigned a value between zero and one. This is a weighting value used to weight a particular criterion's relative importance to the process with respect to the other applicable criteria. The establishment of the weights associated with each selection criterion will be judgmental in nature, but approaches exist that support reliability and consistency in these values. One such approach is the Analytic Hierarchy Process[14] (AHP) which uses pairwise comparisons between the factors to develop the relative weights. The process starts with broad criteria (eg. reproduction of the attrition process) and disaggregates these into component criteria that are much easier for the human mind to compare consistently. When comparison is accomplished on a lower level the process then reassembles the values to establish the relative weights between the criteria in question. This approach is qualitative in nature and is less burdensome to implement from a data requirement point of view, than more quantitative approaches. If a more

[14] The reader is referenced to T.J. Saaty's The Analytic Hierarchy Process, McGraw Hill 1980.

quantitative approach is desired the multiattribute utility approach of Keeney and Raiffa[15] may be used. [Ref.31, p.103]

Note that the criteria and values of w_i for a particular model purpose are independent of the model under investigation. The w_i's become a standard set of criteria, with standard relative weights, for validation of models that fall in the same category by purpose type. This provides for objective, standardized comparison of models. Some criteria belong to more than one set, but will be of different relative importance to the validation process within the different sets. Thus, model purpose has the effect of breaking the available evaluation criteria into two sets: those applicable to that purpose and those that are not. Figure 11 illustrates this characteristic.

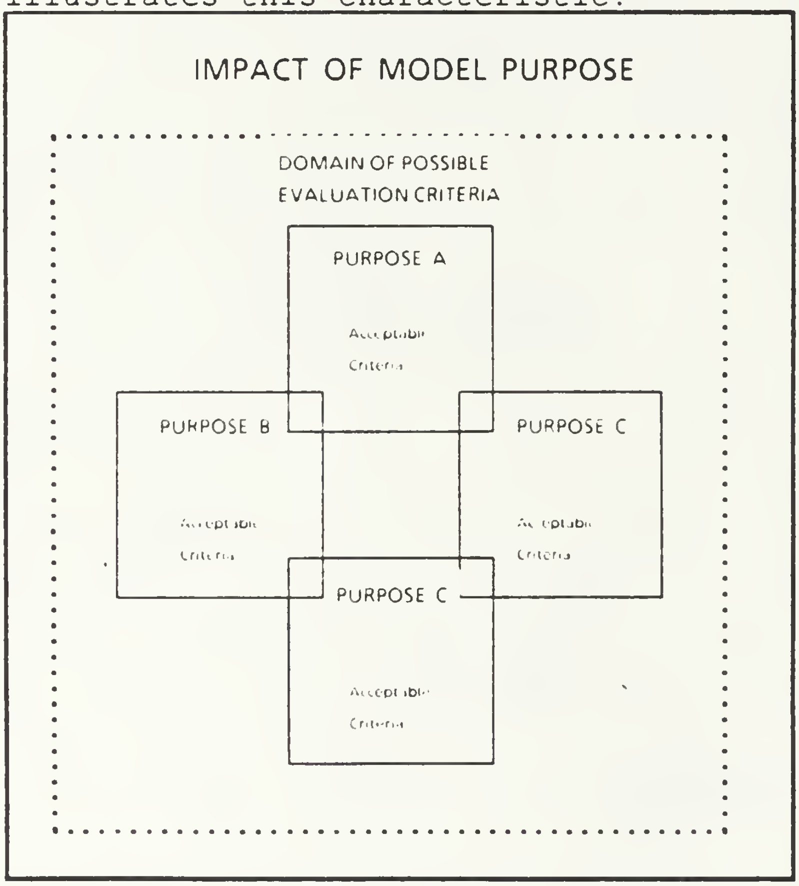

Figure 11

[15] The reader is referenced to Keeney and Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs, Wiley 1976.

The available criteria from which choices can be made are those supported by the NTC data. This data supports a great number of possible criteria and these criteria will grow in number as the data collection efforts at the NTC improve. Definition and enumeration of the currently possible criteria are beyond the scope of this thesis. However, the newly developed technical database is recommended as an appropriate method for maintaining the specific data needed to support the evaluation criteria. Technical databases, in the context of the NTC Research Database, are specifically developed to support particular research efforts, and one of these technical databases could be tailored to support the validation process. Tied directly to the tactical database, the technical database could be set for periodic updates as more data became available. This would provide an automatic method of staying current with the effects of emerging weapon systems and changing tactics and doctrine. As these new weapons and tactics are used at the NTC, the validation database would automatically be updated, reflecting these changes. Validating models and simulations against this type of database would ensure that the models themselves underwent periodic updates, otherwise they would not be validated and therefore not used.

Another important aspect of model purpose is the restriction it places on the comparison of models. Since model validation is the establishment of a particular level of confidence in a model, an obvious extension is the examination of the relative confidence between models. The effect of model purpose is to limit the comparison of models to those in the same purpose category. The validation process, then, operates within the domain of model purpose as illustrated in Figure 12.

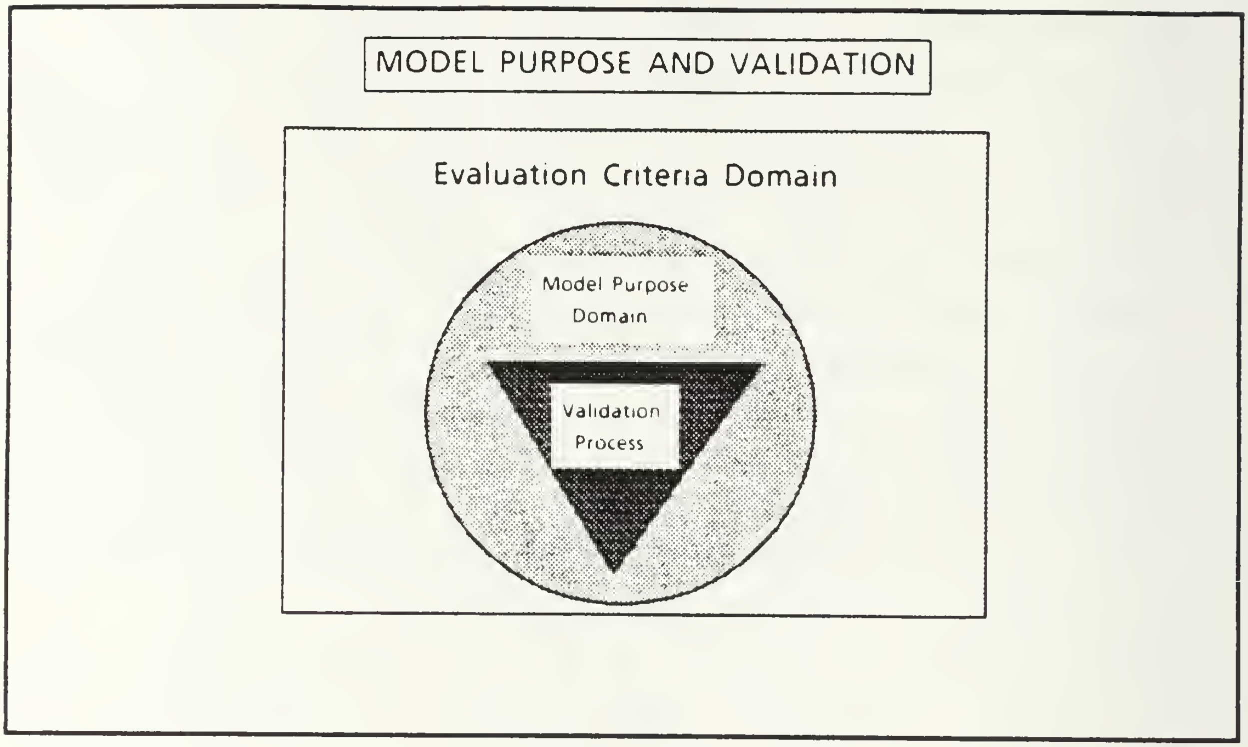

Figure 12

The process of completing step one may be illustrated through considering a specific example. Assume that the set of validation criteria that NTC supports has been defined. CASTFOREM, a high resolution, systemic combat model, is being evaluated and the validation scenario is a standard Soviet Motorized Rifle Regiment attacking an American Mechanized Infantry Battalion. The purpose of the model is to investigate the possible tactical courses of action available to the U.S. commander. The most important tactical aspect of the situation is the maneuver ability of the forces involved (i.e. timeliness and position have a greater impact on the battle than relative weapon characteristics, etc.).

Given the particular scenario and model purpose, a distinct set of applicable validation criteria, A_i, may be

selected from the set of available criteria. Such a selection might consist of the following evaluation criteria.

A_1: Mean loss rate of M1's to T72's in each range band.
1) Range band 1-- < 1000 meters
2) Range band 2-- 1000 meters < rng < 2000 meters
3) Range band 3-- 2000 meters < rng < 3000 meters

A_2: Movement rate per vehicle in each range band.
1) Range band 1-- < 1000 meters
2) Range band 2-- 1000 meters < rng < 2000 meters
3) Range band 3-- 2000 meters < rng < 3000 meters

A_3: Range distribution of M1 shots against T72's.

When these criteria have been selected, relative weights may be assigned to them using one of a number of available methods. These relative weights will be used at a later time, and for the purpose of this example are assumed to be:

$A_1 \text{---}> w_1 = .3$

$A_2 \text{---}> w_2 = .5$

$A_3 \text{---}> w_3 = .3$

At the completion of step 1 the model purpose and scenario have been defined; the evaluation criteria that will be used in the empirical testing portion of the validation process have been identified; and the relative importance of each criterion has been established.

B. ESTABLISH FACE VALIDITY

Establishing Face Validity or the reasonableness of the model is the second stage of the validation process. Those knowledgeable about the real world system being modeled review the model for realism. This is the stage of the validation process where the opinion of experts as well as the lessons of the past can be brought to bear to preclude poor modeling.

The major checks for reasonableness include continuity checks, consistency checks, and response to degenerate and absurd conditions. [Ref.32, p.929]

1. Continuity Checks: small changes in input parameters should cause consequent small changes in output variables unless large changes can be understood and justified by the structure and process of the system being modeled.

2. Consistency Checks: runs of the same scenario should produce similar results, changing the initial seed etc should not produce dramatically different outcomes.

3. Degenerate Conditions: when certain aspects of the model are removed the model output should reflect their absence.

4. Absurd Conditions: absurd conditions should not be generated by the model, e.g. negative counts of things, entities being in more than one place at a time.

The test for face validity has its greatest value early in the modeling process. The model developer should have taken efforts to ensure that checks for reasonableness were accomplished throughout the model building process. In doing this the modeler continually eliminates the more obvious representation errors that the model may contain.

C. EMPIRICAL TESTING

Data from the NTC is used to support steps three and four of the validation process: testing the model's hypotheses and testing the model's predictive capability. These two steps are combined because of their similarity of approach.

The empirical testing of the model involves comparing data from the NTC to data from the model. To accomplish this a portion of the NTC data is generally used to "drive" the simulation. These data are used to ensure that the scenario and other domain characteristics of the NTC and the model are the same. The data required for this are most often that data which reflect the unmodeled human decision processes and that which represent weapon characteristics that are affected by the NTC representation of reality.

The human decisions that most non-systemic high resolution simulations require as input revolve around the maneuver of forces on the battlefield. Thus, the position/location data are often a part of the model "driving" data set and are considered "historical data" of the scenario in question. The second consideration is prompted by the discrepancies that do exist between the MILES representation of a weapon system and the weapon's actual characteristics. The significant weakness of MILES is its

inability to capture the true range of the longer ranged weapons systems. This is overcome by using MILES weapon range characteristics as inputs to the model rather than those officially provided by the Army Material Command. This substitution is acceptable for the validation process if one condition is met. Prior to use of the MILES parameters, the model must be subjected to a sensitivity analysis. This analysis must produce reasonable results over a parameter range that includes both the MILES value and the official value of the parameter in question. This will provide confidence that after successful validation resubstitution of the official values of the parameters will still produce realistic results.

The empirical testing takes place over the range of criteria identified in step one of the validation process. The setting is :

1. An identified set of evaluation criteria
2. A weighting scheme associated with the criteria set
3. NTC data available for two purposes
    a. Provide adequate data to "drive" simulation
    b. Provide adequate data to support comparative evaluation of model data
4. A model producing sufficient data to conduct the test.

A comparison of the data from the simulation and from the NTC will be the eventual test used in the validation process. While validation is essentially a relative process, minimum acceptance levels for each criteria should be established. These bounds should be liberal, giving full consideration for the reliability limitations in a testing process such as this.

Setting the bounds for acceptance regions for most statistical procedures translates into establishing the bounds for acceptable probabilities of Type I and Type II errors. Type I errors, rejecting a valid model, may be considered as the model builder's risk , and Type II errors, accepting am invalid model, may be considered as the model user's risk [Ref.33, p.186]. Generally, minimization of the

Type II error for a specified level of Type I error is the goal of the testing procedure. The probability of a Type I error is referred to as the level of significance associated with the test procedure. The establishment of the level of significance is dependent on two factors:

1) The deviation of model data from NTC data that would be expected if NTC were a perfect surrogate for reality, and

2) The expected deviation of NTC data from reality based on its imperfections as a surrogate.

While consideration of the first factor will generate rather consistent initial values for the levels of significance, the second factor will cause a divergence of values for the various criteria. In the example under consideration, tests for each of the criteria, A_i's, may have the same initial standard for level of significance, say .01. The second factor requires the consideration of the source of data that supports each of the criteria under question. Since the MILES gives less reliable and less accurate data than the position\location system, using the same level of significance to test both criteria would be inappropriate. A model might be improperly rejected based on the additional inaccuracies of the data base, even when it appropriately represents reality. Thus the requirement for a particular level of significance should be relaxed for criteria where the NTC shows significant weakness in representing reality. Relaxing requirements, when speaking of levels of significance, means decreasing the value of the level of significance. This effectively increases the acceptance region of the test. The results of consideration of factor two, for the example, are portrayed in Figure 13. By making these adjustments, the different levels of combat representation that the NTC provides have been accounted for in the testing process.

The next step in this phase of the validation process is to run the simulation and collect data that supports testing of model hypotheses and model predictive capabilities. After

LEVELS OF SIGNIFICANCE

Criteria	Initial (factor 1)	Final (factor 2)
A_1	.10	.01
A_2	.10	.10
A_3	.10	.05

Figure 13

this a comparison of data produced by the NTC and by the model may be accomplished. The theory of statistics, especially in terms of parameter estimation, hypothesis testing and time series analysis provide the tools by which these comparisons may be made. These comparisons result in an acceptance or rejection decision for each of the criteria in question.

The majority of the tests performed will be of the nature of comparing central tendency measures of the identified criteria. If a large enough sample size of both the NTC data and the simulation data is available, then the Central Limit Theorem may be invoked, and a two sample Z-test can be used to compare the sample means of the data collected in support of each criteria. Efforts should be made to ensure large sample sizes because this provides the most straightforward test of the criterion. If this is not possible, the two sample t-test may be applied if the sample populations can be shown to be normal or nearly normal, and the variances of the NTC data and the simulation data can be shown to be approximately the same. If these conditions cannot be established, non-parametric tests may be needed, because of the distribution-free requirements imposed by the data.

If the model fails to pass any of the tests associated with the evaluation criteria, the model should be rejected. The criteria that caused the rejection should be reported to the modeler for corrective action. Those models that pass these tests form the feasible set, from which the decisionmaker may choose a model to apply to the problem at hand.

When the feasible set of models within a particular purpose domain have been established, there remains the process of deciding which model to use. The analyst must present to the decisionmaker the information necessary for choosing a specific model in a succinct, yet meaningful form. One method, certainly not the only method, which gives the decisionmaker both flexibility and advice as to the proper model choice involves P-values[16]. The decisionmaker would be provided two pieces of information for each criterion tested. The first would be the weighting factor initially established in step one. This provides the decisionmaker a basic recommendation of the relative importance of the criteria under question. The decisionmaker, while not obligated to use these specific weights in his decision process, will most probably use these as a baseline upon which to apply refinements. These refinements of the relative importance of each criteria will be based on his personal perceptions of the problem under consideration and account for minor changes in the problem structure that occurred during the validation process. The second piece of information is a vector of the P-values associated with the criteria against which the each model was evaluated. A vector of these values is provided for each of the models in the feasible set. This provides the decisionmaker with information on the margin of acceptance by

---

[16] For those unfamiliar with the idea of P-values, Probability and Statistics for Engineering and the Sciences (Brooks/Cole 1987) by Jay Devore provides a succinct description of their application.

which each model successfully met each requirement. This allows differentiation between models that barely met certain criteria and models that met the criteria by a wide margin.

Through this process of empirical testing the set of candidate models is reduced to a set of feasible models. The decision maker is then provided information to assist him in making an appropriate selection from the feasible set. This process is outlined in Figure 14.

D. SUMMARY

This process and its results have the potential to benefit the Army in many ways. First it provides a method of selecting a model between competing candidates. Second, it will highlight the significant deficiencies of each model put through the process. Finally, it provides an objective alternative to the subjective methods of validation that are predominant in the Army today.

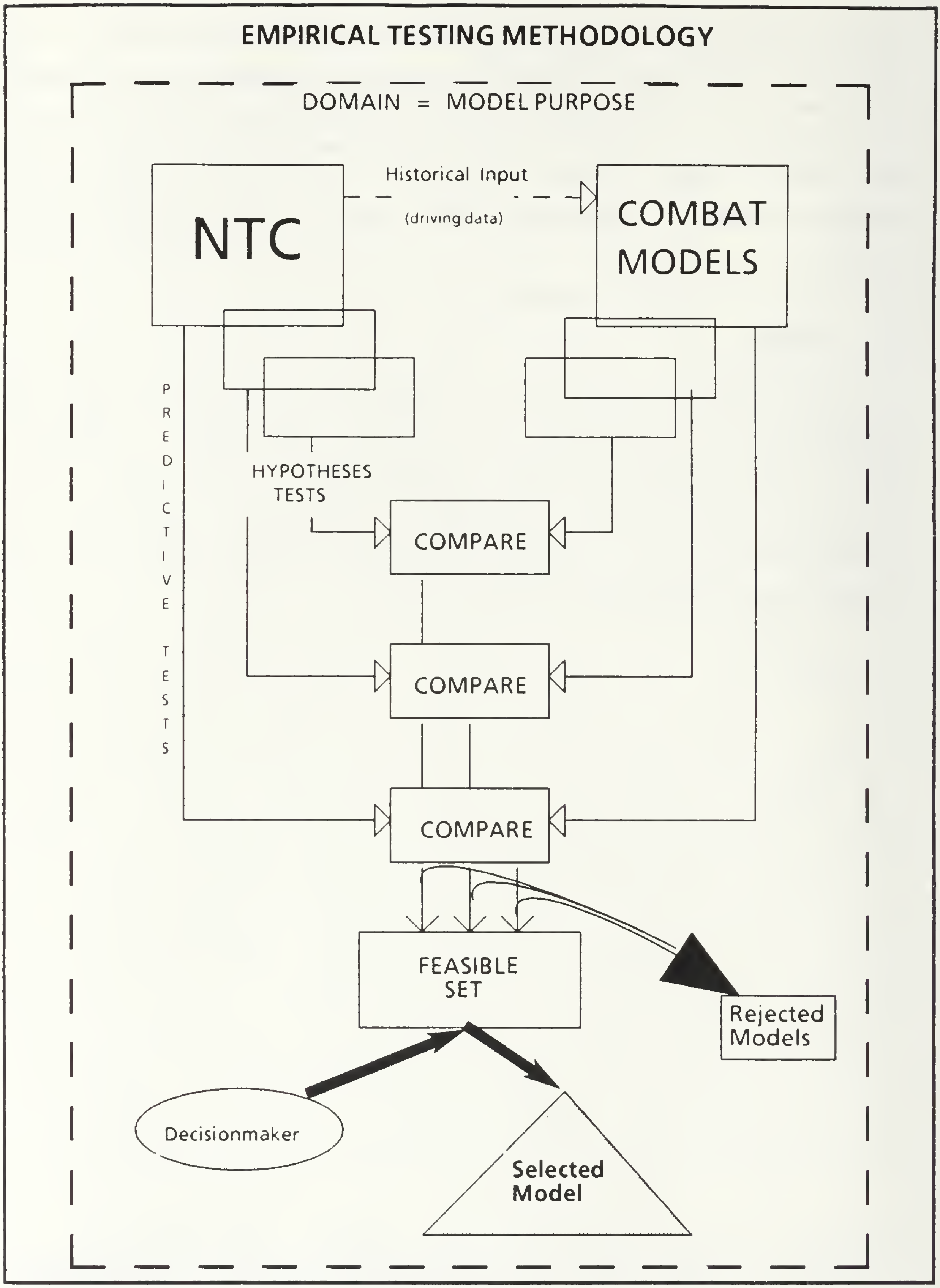
EMPIRICAL TESTING METHODOLOGY
DOMAIN = MODEL PURPOSE
NTC
Historical Input
(driving data)
COMBAT MODELS
PREDICTIVE TESTS
HYPOTHESES TESTS
COMPARE
COMPARE
COMPARE
FEASIBLE SET
Rejected Models
Decisionmaker
Selected Model

Figure 14

VII. CONCLUSION

The National Training Center was found to be a source of data that is highly representative of actual combat. Lack of this kind of data has been a serious hindrance to attempts to validate high resolution combat models. It is recommended that a technical database, under the umbrella of the National Training Center Research Database, be developed and maintained to support efforts to validate combat models.

The methodology presented provides an approach to the issue of validation that makes use of the data from NTC, automatically updates validation criteria to account for changes in weapons and tactics, and is responsive to the purpose for which the model was developed.

APPENDIX A

COGNITIVE BIASES[17]

1. Availability:	The tendency to use only easily available information and ignore less available sources of significant information. An event is believed to occur with high probability if it is easy to recall similar events.
2. Conservatism:	Failure to revise estimates as much as they should be, based on receipt of new significant information.
3. Data Saturation:	Tendency to reach premature conclusions based on a small amount of data, ignoring data received later.
4. Ease of Recall:	Data which can be easily recalled or assessed will affect perception of the likelihood of that event. People typically weigh easily remembered data more than that not so easily remembered.
5. Expectations:	People often remember and attach higher validity to information which confirms their previously held beliefs than they do to disconfirming information.
6. Fact-Value Confusion:	Strongly held values may often be regarded and presented as facts. That type of information is sought that lends credibility to one values and views.
7. Fundamental Attribution Error:	The tendency to associate success with personal ability and failure with poor luck.
8. Gamblers Fallacy:	False assumption that an unexpected occurrence of a "run" of one event enhances the probability of another event occurring.
9. Hindsight:	People are often unable to think objectively if they receive information that an event has occurred and they are told to ignore this information. With hindsight outcomes that have occurred seem to have been inevitable.

[17] This information has been selected from Andrew P. Sage's article "Behavioral and Organizational Considerations in the Design of Information Systems and Processes for Planning and Decision Support. It only represents a limited selection of existing cognitive biases.

10. Illusion of Control: A good outcome in a chance situation may well have resulted from a poor decision. The individual may assume a feeling of control over events that is not reasonable.

11. Illusion of Correlation: Mistaken belief that two events covary when they do not.

12. Law of Small Numbers: Lack of sensitivity to quality of evidence. Tendency to put greater confidence in predictions based on small samples of data with nondisconfirming evidence than in much larger samples with minor disconfirming evidence. Sample size and reliability often have little effect on relative confidence.

13. Order Effects: The order in which information is presented affects information retention in memory.

14. Redundancy: The more redundant the data, the more confidence associated with it, even if it is the same data presented in different ways.

15. Regression Effects: The largest observed values of observations are used without regressing towards the mean to consider the effects of noisy measurements. Tendency to ignore uncertainties.

16. Selective Perceptions: The tendency to select from the information available only that information that conforms to already held views.

17. Spurious Cues: Often cues appear only by occurrence of a low probability event but are accepted as commonly occurring.

APPENDIX B[18]

MULTIPLE INTEGRATED LASER ENGAGEMENT SYSTEM

MILES simulates the fire of direct fire weapons systems and is used for engagement simulation. It consists of a receiver-transmitter combination which uses eye-safe Gallium Arsenide lasers to simulate the fire of direct fire weapon systems. The MILES transmitter is a coded beam laser transmitter which is attached to the weapon whose fire it is simulating. Within MILES, a complete hierarchy of weapons from the M16 to the TOW missile is made available through beam coding. By coding the beam, being able to measure its intensity, and using logic circuits in the receiver, MILES is able to enforce proper engagement techniques for particular weapon systems and to provide realistic operating ranges and hit/kill probabilities. The MILES transmitter is sound-activated, sending its coded beam only when a blank from the weapon is actually fired, thus forcing logistical play and requiring weapons to be operational. If blanks are not available for a particular weapon system Miles may be adapted to fire without blanks. In this mode, the transmitter employs a logic circuit which counts the number of rounds expended and enforces a mandatory reload point for larger systems such as the TOW or Dragon. When the basic load is expended, the transmitter is disabled and requires resetting before the weapon can fire again. Controllers reset the transmitter once the requirements of resupply have been met.

The MILES receiver works with a group of laser detectors that are attached at prominent places on individual soldiers and vehicles. When the coded laser pulses are received from

---

[18] This description of the MILES is, with minor modifications, from the excellent discussion given by Timothy Reischl in Reference .

a transmitter, the received codes are analyzed by the receiver. The arriving pulses are compared to a threshold level. If the pulse strength exceeds the threshold, the weapon is in range, and a single bit is registered in the detection logic. Once a valid arrangement of bits is formed corresponding to a code for a particular weapon, a decision is made to determine "hit" or "near miss". To accomplish a relative difference in the probability of "hit" to "near miss", MILES uses two approaches. First the transmitter emits a smaller number of "hit" messages than "near miss" messages, giving a lower probability of hit than near miss. Secondly, the transmitter operates at higher power when it emits near miss messages, thus increasing the area over which near misses will be recorded.

Once a "hit" is registered, the receiver, reading from the codes on the beam, determines if the firing weapon can kill the vehicle carrying the receiver. (This precludes the "killing" of tanks with an M16.) If this is the case, it determines the extent of damage to the vehicle. The receiver then causes audio and visual signal to be sent of the individual of crew to announce the hit or near miss. The kill indications are a flashing strobe light for vehicles and a distinctive, continuous beeping noise for personnel. When a kill occurs, the "killed" weapon is disabled from further use.

Through these methods MILES provides an extremely realistic simulation of weapon firing and the casualty effects of weapons engagements.

APPENDIX C

NTC TACTICAL DATABASE TABLES[19]

1. Mission Identification Table

Purpose: To provide information to completely identify and categorize each mission segment.

Data Elements:

Mission start date and time
Mission end date and time
History Name
Segment Number
Mission Type
Unit ID
A (armored) or M (mechanized)

2. Player State Initialization Table

Purpose: To describe the participants at the beginning of the mission segment. Includes friendly, enemy and controllers.

Data Elements:

Player Identification (vehicle bumper #)
Logical Player Number
B (blue), O (opfor), or W (white)
I (instrumented) or N (not instrumented)
Player Type Code
Next Higher Line Unit
T (tracked) or U (untracked)
Player Status Code

3. Player State Update Table

Purpose: To track changes to all participants throughout the mission segment.

Data Elements:

Date and Time of Update
Player identification (vehicle bumper #)
Logical Player Number
B (blue), O (opfor), or W (white)
I (instrumented) or N (not instrumented)
Vehicle Type Code
Next Higher Line Unit
T (tracked) or U (untracked)
Player Status Code

---

[19] A separate INGRESS database consisting of this set of tables is created for each mission

4. Unit State Initialization Table

Purpose: To describe Opfor and Bluefor units at the beginning of the mission segment.

Data Elements:
- Unit Name
- Next Higher Line Unit
- Next Higher Statistical Unit
- Unit Type Code
- Force Code (R or B)
- Echelon

5. Unit Type Table

Purpose: To provide information relating to unit organizations.

Data Elements:
- Unit Type
- Unit Force (R or B)
- Echelon Identifier
- Unit Description

6. Unit State Update Table

Purpose: To track changes to all units throughout the mission segment.

Data Elements:
- Date and Time of Update
- Unit Name
- Next Higher Statistical Unit
- Unit Type Code

7. Player/ Vehicle/ Weapon Code Table

Purpose: To define a unique code for each weapon on the battlefield. The codes are the same as the MILES codes.

Data Elements:
- Side Code (R or B)
- Player Type Code
- Vehicle Description
- MILES Weapon Code
- Weapon Description
- Initial Ammunition Load

8. Firing Event Table

Purpose: To maintain a time ordered record of all legitimate firings recorded by the RDMS.

Data Elements:
- Date and Time of Fire Event
- Player ID
- Logical Player Number
- MILES Weapon Code
- Position Location X Coordinate
- Position Location Y Coordinate
- Ammunition Remaining

9. Pairing Event Table

Purpose: To maintain a time ordered record of legitimate pairing events. Includes information on firer if the pairing event can be matched with a fire event.

Data Elements:
- Date and Time of Pairing
- Target ID
- Target LPN
- N (near miss), H (hit), K (kill)
- Firer Weapon Type (MILES)
- Fratricide Indicator (Y/N)
- Target Position Location X Coordinate
- Target Position Location Y Coordinate
- Firer Position Location X Coordinate
- Firer Position Location Y Coordinate

10. Communication Table

Purpose: To maintain a record of all commo events. Tracks key depressions and releases by mission segment.

Data Elements:
- Date and Time of Commo Event
- Player ID
- LPN
- Radio Net (1 or 2)
- Duration of Transmission (sec)

11. Ground Player Position Location Table

Purpose: To maintain a time-ordered record of Position location for each instrumented ground participant. Can be recorded at selected intervals.

Data Elements:
- Date and Time of Position Location
- Player ID
- LPN
- Position Location X Coordinate
- Position Location Y Coordinate

12. Air Player Position Location Table

Purpose: To maintain a time-ordered record of Position location for each instrumented air player. Can be recorded at selected intervals.

Data Elements:
- Date and Time of Position Location
- Player Id
- LPN
- Position Location X Coordinate
- Position Location Y Coordinate
- Position Location Z Coordinate

13. Indirect Fire Casualty Assessment (IFCAS) Target Table

Purpose: To maintain a list of pre-planned indirect fire targets and their locations.

Data Elements:
- IFCAS Target Name
- Side (R or B)
- Target Index
- Position Location X Coordinate
- Position Location Y Coordinate

14. IFCAS Target Group Table

Purpose: To maintain a list of pre-planned IFCAS target groups and their component targets.

Data Elements:
- IFCAS target Group Name
- Side (R or B)
- IFCAS Target Name #1
- IFCAS Target Name #2

(Up to 10 targets)

15. IFCAS Missions Fired Table

Purpose: To maintain a list of all IFCAS missions fired during this mission segment.

Data Elements:
- Date and Time of IFCAS Mission
- IFCAS Preplanned Mission Number
- Force Code (R or B)
- Battery Identification
- IFCAS Target Group Name
- IFCAS Target X Coordinate
- IFCAS Target Y Coordinate
- IFCAS Weapon Type Code
- Shell Type Code
- Fuse Type Code

16. IFCAS Casualties Table

Purpose: To maintain a list of all casualties assessed as a result of IFCAS missions fired during mission segment.

Data Elements:
- Date and Time of IFCAS mission
- IFCAS Mission ID
- Force Code (R or B)
- ID of Player Killed by IFCAS
- LPN of Player Killed by IFCAS
- Target Position Location X Coordinate
- Target Position Location Y Coordinate

17. Minefield Casualties Table

Purpose: To maintain a list of all casualties assessed as a result of minefields during mission segment.

Data Elements:
- Date and Time of Minefield Casualty
- ID of Player Killed by Minefield
- LPN of Player Killed by Minefield
- Target Position Location X Coordinate
- Target Position Location Y Coordinate

18. Control Measure Table

Purpose: To maintain a list of all control measures established at the beginning of mission segment.

Data Elements:
- 1: Blue 2: Opfor
- Operating System Code
  - 0: Maneuver
  - 1: Fire Support
  - 2: Intelligence
  - 3: Mobility / Counter Mobility
  - 4: Communications
  - 5: Air Defense
  - 6: Unspecified
- Echelon Code
  - 0: Platoon
  - 1: Company
  - 2: Battalion
  - 3: Regiment/ Brigade
  - 4: Division
- Type: 1=Point, 2=Line, 3=Area
- Purpose
- Mine Type (if applicable)
- Number of Points Used
- X Coordinate, Point 1
- Y Coordinate, Point 1
- X Coordinate, Point 2
- Y Coordinate, Point 2

(Up to 12 Points)

19. Control Measure Add Table

Purpose: To maintain a list of all control measures added during mission segment.

Data Elements:
- 1: Blue 2: Opfor
- Operating System Code
  - 0: Maneuver
  - 1: Fire Support
  - 2: Intelligence
  - 3: Mobility / Counter Mobility
  - 4: Communications
  - 5: Air Defense
  - 6: Unspecified

Echelon Code
0: Platoon
1: Company
2: Battalion
3: Regiment/ Brigade
4: Division

Type: 1=Point, 2=Line, 3=Area

Purpose

Mine Type (if applicable)

Number of Points Used

X Coordinate, Point 1

Y Coordinate, Point 1

X Coordinate, Point 2

Y Coordinate, Point 2

|

(Up to 12 Points)

REFERENCE LIST

1. Airpower Research Institute Report AU-ARI-84-8, A Conceptual Design For A Model To Meet The War-Gaming Needs Of The Major Commands Of The United States Air Force, by D. B. Fox, p.29, July 1985.

2. Flanagan, S. A., "Verification and Validation: A TRAC Approach", Phalanx, v. 21, p. 21, March 1988.

3. Center For Naval Analyses (CNA), CNA RC 115, Warfare Model Verification Conference, p. 4, 1 October 1968.

4. Thomas, Clayton J., "Verification Revisited", Military Modeling, edited by Wayne P Hughes, Military Operations Research Society, p. 298, 1984.

5. Shannon, Robert E., "Tests for the Verification and Validation of Computer Simulation Models", 1981 Winter Simulation Conference Proceedings, IEEE, 1984.

6. Encyclopedia Britannica, Vol K, p.419.

7. Ayer, A. J., Logical Positivism, The Free Press, 1959.

8. Thomas, Clayton J., "Verification Revisited -- 1983", Military Modeling, edited by Wayne P. Hughes, p.309, Military Operations Research Society, 1984.

9. Blaug, M., Economic Theory in Retrospect, p. 612, Richard D. Irwin, 1962.

10. Forrester, J. W., Urban Dynamics, p. 144, Massachusetts Institute of Technology Press, 1969.

11. Naylor, Thomas and Finger, J. M., "Verification of Computer Simulation Models", Management Science, v.14, p. B-92, October 1967.

12. Reichenbach, Hans, The Rise of Scientific Philosophy, p. 304, University of California Press, 1951.

13. Gass, Saul and Thompson, Bruce, "Guidelines for Model Evaluation: An Abridged Version of the U.S. General Accounting Office Exposure Draft", Operations Research, v.28, p.445, March-April 1980.

14. Schrank, William E. and Holt, Charles C., "Critique of: Verification of Simulation Models", Management Science, v.14, p. B-105, October 1967.

15. Hermann, Charles F., "Validation Problems in Games and Simulations With Special Reference to Models of International Politics", Behavioral Science, v.12, p. 219, 1967.

16. House, Peter W., "Diogenes Revisited - The Search for a Valid Model", Simulation, p.122, October 1984.

17. Sage, Andrew P., "Behavioral and Organizational Considerations in the Design of Information Systems and Processes for Planning and Decision Support", IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-11, p. 646, September 1981.

18. Class Notes, OS3636 Architecture of C3I Systems, Professor James G. Taylor, 1987.

19. English, John A., A Perspective on Infantry, p. vii, Praeger Publishers, 1981.

20. Liebholz, Stephen W., "Twenty Questions", Military Modeling, edited by Wayne P. Hughes, p. 646, Operations Research Society, 1984.

21. Mcquie, Robert, "Military History and Mathematical Analysis", Military Review, v.5, p. 10, U.S. Command and General Staff College, 1970.

22. Rand Note N-2461-A, Utilizing the Data from the Army's National Training Center: Analytical Plan, by R. A. Levine, J. S. Hodges, and M. Goldsmith, p. 1, June 1986.

23. Rand Note 87-02, Observations from Three Years at the National Training Center, by L. E. Word and C. A. Johnson, p.19, January 1987.

24. Fobes, J. L., National Training Center Handbook, p. v, U.S. Army Research Institute, November 1983.

25. Curry, Thomas P., "Are We Training to Fight that First Battle ... and Lose?", Armor, v.90, p. 20, July-August 1981.

26. ARI Research Note 87-08, Capability to Analyze National Training Center Data, by J. Briscoe, p. 2, January 1987.

27. Galing, B.W. and Wimberly, R. J., Realism in Combat Models, paper presented at MORS Conference in Washington D.C., April 1988.

28. Reischl, T. J., An Examination of Battalion Training at the National Training Center, M.S. Thesis, Naval Postgraduate School, Monterey, California, March 1980.

29. ARI Research Note 87-15, National Training Center Lessons Learned: Data Requirements, by C. A. Johnson, p.8, February 1987.

30. ARI Research Note 511C1, NTC Tactical Database Preliminary Design (Revised), by J. A. Briscoe and J. D. Baldwin, p.2, April 1987.

31. Bard, Jonathan F., "Evaluating Space Station Applications of Automation and Robotics", IEEE Transactions on Engineering Management, v.EM-33, pp.103-104, May 1986.

32. Schlesinger, J.R., and others, "Developing Standard Procedure for Simulation, Validation, and Verification", Proceedings of 1974 Summer Conference, pp. 927-933.

33. Balci, O. and Sargent, R. G., "Validation of Multivariate Response Models Using Hotelling's Two-sample T^2 Test", Simulation, p.186, December 1982.

INITIAL DISTRIBUTION LIST

		no. copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Professor Sam H. Parry, Code 55Py Department of Operations Research Naval Postgraduate School Monterey, CA 93943-5000	2
4.	Major Bernard W. Galing TRAC Research Element Monterey, CA 93943-5000	1
5.	Professor Carl Jones, Code 74 C3 Academic Group Naval Postgraduate School Monterey, CA 93943-5000	1
6.	Professor Peter Purdue, Code 55Pd Department of Operations Research Naval Postgraduate School Monterey, CA 93943-5000	1
7.	Curricular Office, Code 39 Naval Postgraduate School Monterey, CA 93943-5000	1
8.	Curricular Office, Code 55 Naval Postgraduate School Monterey, CA 93943-5000	1
9.	Deputy Undersecretary of the Army for Operations Research Room 2E261, Pentagon Washington D.C. 20310	1
10.	Mr Gerald K. Coville RFD 1 South Main St Stockholm, Me 04783	1
11.	CPT Michael P. Coville RFD 1 South Main St Stockholm, Main 04783	1

Thesis
C75686 Coville
c.1 A methodology for
validation of high reso-
lution combat models.